THE APPLICATION OF FOURIER TRANSFORM NEAR INFRARED (FT-NIR) SPECTROSCOPY IN THE WINE, FRUIT AND DRIED FRUIT INDUSTRIES OF SOUTH AFRICA

ANINA VAN ZYL

Thesis presented in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE IN FOOD SCIENCE

Department of Food Science
Faculty of Agricultural and Forestry Sciences
University of Stellenbosch

Study Leader: Dr. M. Manley
Co-study Leader: Dr. E. Joubert
Co-study Leader: Mr. E.E.H. Wolf

September 2000

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

ABSTRACT

This study shows Fourier transform near infrared (FT-NIR) spectroscopy's application on wine, fruit and dried fruit for quantitative determinations or as a discriminative method for classification purposes.

During wine production optimum yeast growth, resulting in healthy alcohol fermentation rates, is monitored by the amount of free amino nitrogen (FAN) present in the must. The status of malolactic fermentation (MLF) in Chardonnay wines is monitored by determining the degree of conversion of malic to lactic acid. Ethyl carbamate (EC), a suspected carcinogen, is mainly formed during ageing of wine and is restricted by legislation in some countries. It is therefore necessary to determine the EC content in wine.

Fourier transform near infrared spectroscopy can be used on wine as a rapid method to measure the °Brix (residual sugars) content of must and to discriminate between different must samples in terms of their FAN values. It can also be used as a rapid method to discriminate between Chardonnay wine samples in terms of the MLF status and between table wine samples in terms of the EC content. Calibrations were derived and it was found that a very strong correlation existed in the sample set for the FT-NIR spectroscopic predictions for °Brix ($r = 0.99$, SEP = 0.31%), but poorer correlations for the FAN ($r = 0.405$, SEP = 275%), malic acid ($r = 0.64$, SEP = 1.02%), lactic acid ($r = 0.61$, SEP = 1.35%) and EC predictions ($r = 0.47$, SEP = 3.6%). When soft independent modelling by class analogy (SIMCA) was applied as a discriminative method, the must and wine samples were classified in terms of their FAN and EC values and MLF status, respectively, obtaining results with recognition rates exceeding 80%.

The canning of peaches has become a significant industry in South Africa, producing approximately 4.1 million cartons per year. Fourier transform near infrared spectroscopy was applied as an alternative non-destructive method for the quantitative determination of the total soluble solid (TSS) content of whole fresh peaches. The TSS content of fresh clingstone peaches is an indication of the internal quality, maturity and perceived sweetness of the fruit for the peach canning industry.

By determining the TSS, fresh peaches can be graded and the farmers compensated accordingly. Results obtained by building QUANT+^{TM} calibrations for the TSS content (r = 0.96, SEP = 0.55%) showed acceptable accuracy and can replace the present destructive methods.

Peak periods during the harvesting season necessitate storage of peaches for up to three weeks before canning. Approximately 5 - 10% of the peaches stored, disintegrate during canning due to loose skin, large stone cavities, soft flesh and rot. The storage potential of fresh clingstone peaches can be successfully predicted with FT-NIR and SIMCA models, using subjective internal quality evaluations. Results with recognition rates exceeding 80% were obtained in most cases and this method proved useful as a non-destructive method of quality assessment. By applying this method, losses caused when storing peaches with poor storage quality will be reduced.

The golden sultana industry plays an important role in the dried fruit exporting market of South Africa. Due to the large numbers of consignments that must be checked upon arrival, and the need for rapid decision making during processing, it is essential to replace the present time-consuming analytical methods. Fourier transform near infrared spectroscopy was used as a rapid, analytical technique to determine whether the SO_2 and moisture contents of sultanas are within specifications upon arrival at the factory and during processing. High positive correlation was found between the measured values and those predicted by FT-NIR spectroscopy for SO_2 (r = 0.99, SEP = 24.09%) and moisture (r = 0.99, SEP = 0.051%) contents.

UITTREKSEL

Hierdie studie dui op Fourier transformasie naby-infrarooi (FT-NIR) spektroskopie se toepassing op wyn, vrugte en droëvrugte vir die uitvoer van kwantitatiewe bepalings of vir klassifikasiedoeleindes om as 'n diskriminasiemetode te dien.

Gedurende wynproduksie word die optimum groei van giste wat lei tot 'n gesonde alkohol fermentasie gemonitor deur die hoeveelheid vry-aminostikstof (VAS) wat in die mos teenwoordig is te bepaal. Die status van appelmelksuurgisting (AMG) in Chardonnay wyne word gemonitor deur die mate van omskakeling van appelsuur na melksuur te bepaal. Etielkarbamaat (EK), 'n vermoede karsinogeen wat hoofsaaklik in verouderde wyne voorkom, word in sekere lande deur wetgewing beperk en dus die bepaling van die EK inhoud van wyne noodsaak.

Fourier transformasie naby-infrarooi spektroskopie kan op mos toegepas word as 'n vinnige metode vir die bepaling van °Brix (residuele suiker) en om tussen die monsters te onderskei in terme van hulle VAS inhoud. FT-NIR kan ook gebuik word as 'n vinnige metode om tussen Chardonnay monsters te onderskei op grond van die status van AMG en tussen tafelwyn monsters op grond van die EK inhoud. Kalibrasies is ontwikkel en daar is gevind dat baie sterk korrelasies bestaan in die monsterstel vir die FT-NIR spektroskopiese voorspelling van °Brix ($r = 0.99$, SEP = 0.31%), maar swakker korrelasies vir die VAS ($r = 0.405$, SEP = 275%), appelsuur ($r = 0.64$, SEP = 1.02%), melksuur ($r = 0.61$, SEP = 1.35%) en EK voorspellings ($r = 0.47$, SEP = 3.6%). Met die toepassing van *soft independent modelling by class analogy* (SIMCA) as diskriminasie metode, is die mos- en wynmonsters geklassifiseer op grond van hul VAS en EK waardes en die status van AMG, en herkenningswaardes van bo 80% is onderskeidelik behaal.

Die inmaak van perskes het 'n beduidende industrie in Suid-Afrika geword en produseer jaarliks ongeveer 4.1 miljoen kartonne. Fourier transformasie naby-infrarooi spektroskopie is toegepas as alternatiewe, nie-beskadigende metode om kwantitatiewe bepalings van die totale oplosbare vastestowwe (TOV) inhoud van heel vars perskes, te doen. Vir die perske inmaak-industrie is die TOV inhoud van vars taaipitperskes 'n aanduiding van interne kwaliteit, rypheid en die soetheid van die

vrugte. Vars perskes kan gradeer word deur die TOS te bepaal en sodoende kan boere oreenkomstige vergoeding ontvang. Resultate wat verkry is deur QUANT+^{TM} kalibrasies vir TOS inhoud te ontwikkel (r = 0.96, SEP = 0.55%), dui op aanvaarbare akkuraatheid en kan die huidige metodes vervang.

Tydens oestyd kom piektye voor wanneer dit soms nodig is om perskes vir tot drie weke op te berg voordat dit ingemaak kan word. Ongeveer 5 tot 15% van hierdie opgebergte perskes disintegreer tydens inmaak omdat opberging lei tot defekte in die perskes soos skille wat loskom, groot pitholtes, sagte vleis en vrot. Die opbergingspotentiaal van vars taaipitperskes kan suksesvol voorspel word deur FT-NIR en SIMCA modelle te bou en subjektiewe interne kwaliteitsevaluerings daarop toe te pas. Herkenningsresultate wat 80% in die meeste gevalle oorskry, is behaal wat hierdie metode as 'n suksesvolle nie-beskadigende kwaliteitbepalingsmetode bewys. Hierdie metode sal verliese kan beperk wat voorkom as gevolg van opberging van perskes met swak opbergingskwaliteit.

Die goue sultana industrie speel 'n belangrike rol in die droë vrugte uitvoermark van Suid-Afrika. As gevolg van die hoeveelheid monsters wat gelyktydig getoets moet word en besluite wat vinnig geneem moet word tydens prosessering, is dit belangrik om die huidige tydrowende analitiese metodes te vervang. Fourier transformasie naby-infrarooi spektroskopie is gebuik as 'n vinnige, analitiese tegniek om tydens ontvangs by die fabriek en gedurende prosessering te bepaal of die SO_2- en voginhoud van goue sultanas binne die spesifikasies val. Goeie positiewe korrelasie is gevind tussen die bepaalde en voorspelde FT-NIR spektroskopiese waardes vir SO_2- (r = 0.99, SEP = 24.09%) en voginhoud (r = 0.99, SEP = 0.051%).

CONTENTS

Language and style used in this thesis are in accordance with the requirements of the *International Journal of Food Science and Technology*. This thesis represents a compilation of manuscripts where each chapter is an individual entity and some repetition between chapters has, therefore, been unavoidable.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following persons and institutions that have formed an integral part of my research study:

Dr Marena Manley, my study leader and senior lecturer at the Department of Food Science, University of Stellenbosch, for her consistent enthusiasm, initiative and guidance from the preparation to the completion of this thesis;

Dr Lizette Joubert, my co-study leader, for her unconditional help, hard work and confidence in my research projects;

Mr Erhard Wolf, my other co-study leader, for his involvement with this project, extra time and work put into this project and his patience with all my questions;

Prof T.J. Britz, for his contribution towards my studies as Chairman of the Department of Food Science, University of Stellenbosch;

Mrs Marianne Reeves, for her help with administrative work and computer assistance;

Mr Marais Blom, Mr Ludick Arnolds & Ms Corine Moes at Distillers and Ms Caroline Steger at the Department of Wine Biotechnology, US, for their support, advice and assistance in supplying wine and must samples and chemical analyses values;

Mr Michael Waldner, Mrs Ewarda Swart en Dr Johann Marais of the ARC Infruitec-Nietvoorbij, for their support in providing lots of valuable information as well as samples and chemical analyses values;

Mr George Dico, ARC Infruitec-Nietvoorbij, for his great help in doing SO_2 and moisture determinations;

Ms Ester Lotz, ARC Infruitec-Nietvoorbij, for her involvement with the peach project and for sharing valuable knowledge;

Mr Dappies Smit, Dried Fruit Technical Services, for providing necessary information on request;

Mr Johann Marais, S.A. Dried Fruit Co-operative LTD, for sharing information;

Mr Grame Menzies, Langeberg Technical Services, for sharing information;

Mr Wiehan Victor, Canning Fruit Producers' Association, for sharing information and lending financial support;

The Council for Near Infrared Spectroscopy, for providing the financial support to attend the 10th International Diffuse Reflectance Conference in Chambersburg, Pennsylvania, USA in August, 2000;

Stellenbosch 2000 Merit Bursary, for financial support throughout my post-graduate studies;

Ms Louise van Zyl, for her assistance in NIR spectroscopy procedures as well as her interest and support during my research;

My fellow post-graduate students, who are all my friends, for their help, support and sharing well-deserved coffee breaks;

My family and close friends, for their encouragement and understanding throughout my studies; and

Wickus, who is always there for me.

"The grace of our Lord Jesus Christ be with you all. Amen."

CHAPTER 1

INTRODUCTION

On the 2nd of February, 1659, Jan van Riebeeck, the first commander of the Cape of Good Hope, wrote in his diary: "Today, praise be to God, wine was made for the first time from Cape grapes" (Hands & Hughes, 1997). Since that day, the South African wine industry has come a long way to produce wines of the highest standard and quality.

South Africa's wine industry has 104 179 ha under vines (wine grape varieties only) and a very strong export market with nearly 130 million liters of natural wine exported in 1999, resulting in estimated earnings of R1.541 billion (Y. van der Merwe, South African Wine Industry Statistics, personal communication). Moving slightly inland to the central part of the Western Cape province, the economically significant peach canning industry of South Africa with a production of ca. 150 000 tons of fresh peaches resulted in an annual producers' earnings of R120 million in 1999 (W. Victor, Canning Fruit Producers' Association, personal communication). The dried fruit industry is another industry in South Africa that is relevant as a job creator in a country where unemployment is a nationwide concern. Annually 8500 tons of golden sultanas are produced with an export value of R85 million in 1999 (D. Smit, Dried Fruit Technical Services, personal communication).

In South Africa, many determinations executed during the winemaking process make use of expensive, quantitative, time-consuming analytical methods. It would, however, often be adequate to use a discriminate screening method as these samples need only to be classified as belonging to a certain class or having reached a specified cut-off point or not. There is a need for a method to discriminate between must and wine samples in terms of their free amino nitrogen (FAN) values, the status of the malolactic fermentation and the level of ethyl carbamate (EC) present. Fourier transform near infrared (FT-NIR) spectroscopy can be used as a rapid method to discriminate between different must or wine samples. If the samples are spectroscopically dissimilar, spectral

differences can be used (Perkin Elmer, 1998). If the spectra are similar, sophisticated techniques such as soft independent modelling by class analogy (SIMCA) can be applied.

A scientific quality assessment of fresh peaches prior to canning has not been executed in South Africa, as yet. The total soluble solid content (TSS) (expressed as °Brix) of fresh peaches is an indication of internal quality, maturity and perceived sweetness and has to be determined at a cannery to comply with regulations and to compensate the farmers accordingly. The conventional method for TSS determination is by use of a refractometer that requires maceration of the sample prior to analyses and a definite need arises for an alternative non-destructive analytical method. Another motivation behind the development of a new method is the possibility of on-line sorting based on quality and functional (°Brix content) parameters.

During the harvesting season's peak periods, storage of peaches for up to three weeks before canning is necessary and during that time approximately 5 - 15% of the peaches stored, disintegrate due to poor internal quality. This is not necessarily linked to maturity. A prediction model for evaluation of the storage potential of canning peaches will reduce subsequent losses caused by storing, as those peaches with poor storage quality could be canned as soon as possible after harvesting. When predicting the storage potential of peaches, they are inspected externally or cut in half for internal inspection, all based on the assessor's experience. Non-destructive FT-NIR could be used as an objective classification method to discriminate between the peaches based on their internal quality and maturity and thus storage potential.

Similar to the wine industry, the golden sultana industry is in need of a method to replace present time-consuming analytical methods used. The moisture content of golden sultanas is determined upon delivery to the storage depot during the harvesting season to ensure that values are below 14%. The moisture and SO_2 contents are also determined during processing for local and international markets as part of quality control. The methods currently in use need to be replaced as the large numbers of consignments that must be checked

upon arrival and rapid decisionmaking during processing, is a problem for the industry. Similarly, FT-NIR spectroscopy can be used as a rapid, and potentially non-destructive analytical technique to determine whether the SO_2 and moisture contents of golden sultanas are within specifications upon arrival at the factory and during processing.

The aims of this study were to:

- apply SIMCA on FT-NIR spectra of must and wine samples to discriminate between the samples in terms of their FAN values, the status of the malolactic fermentation and the level of EC present;
- develop a non-destructive FT-NIR method to replace the present method for determination of total soluble solid content (TSS) of clingstone peaches;
- develop an objective SIMCA FT-NIR method based on subjective internal quality parameters to classify fresh clingstone peaches on reception at a cannery into classes, indicating their storage potential; and
- develop a rapid FT-NIR method to replace the present time-consuming and labour intensive analytical laboratory methods for determination of SO_2 and moisture contents of golden sultanas.

References

Hands, P. & Hughes, D. (1997). *Wines and brandies of the Cape of Good Hope: The definitive guide to the South African wine industry*. P. 13. Somerset West, South Africa: Stephan Phillips Publishers.

Perkin Elmer (1998). *FT-NIR spectroscopy application notes*. The Perkin Elmer Corporation. Beaconsfield, United Kingdom.

CHAPTER 2
CONTENTS

CHAPTER 2

LITERATURE REVIEW

"Uncertainty is what we fear most and yet uncertainty is the very stuff of life. NIR grew up in this age of uncertainty and is of itself a product of statistical uncertainty and quantum process."

(Murray, 1999)

A. Introduction to near infrared (NIR) spectroscopy

In this literature review the essence of near infrared (NIR) spectroscopy, as referred to by Ian Murray as the quantum processes and statistical analysis, will be explained in such a way that it will hopefully prevent (further) uncertainty.

According to Murray (1999) and Norris (1999) NIR spectroscopy had come a long way since its first application in the 1960's which is clear when considering the NIR giants' memories about NIR spectroscopy in their earlier years of research. In those early years several restrictions had to be accepted on the range of tests and the number of samples that could be tested. Only a few favourite wavelengths they knew would work were available using the instruments of that time. As the potential of NIR became known, time and effort was applied to develop the technology and new methods of making spectral measurements were adapted to this spectral region. Both Ian Murray and Karl Norris admit that all of these developments would not be possible without the tremendous development in computer technology. Nowadays, fast computers, new algorithms and elegant software can be used to transform and unravel spectra as never before (Murray, 1999; Norris, 1999).

The basis of NIR spectroscopy lies within the basic principles of spectroscopy. As Penner (1994) puts it: *"spectroscopy deals with the production, measurement and interpretation of spectra arising from the interaction of electromagnetic radiation with matter"*.

B. Spectroscopic properties

B1. Production of NIR spectra

Energy associated with a ray of light is not distributed continuously through space along the wave's associated electric and magnetic fields, but is rather concentrated in discrete packets called photons (Osborne *et al.*, 1993; Penner, 1994). Furthermore, the energy content of matter is quantized and a set of available energy levels for any given atom or molecule will be distinct for that species. Analytical absorption spectroscopy methods are concerned with measuring the relative amounts of radiant energy absorbed at each frequency. Since different materials absorb at different frequencies these measurements will determine the amount of these various substances in a mixture, thus leaving a distinct fingerprint of the material.

Karl Norris (1989), the father of NIR spectroscopy defined the NIR spectroscopy method of analysis as "*an instrumental method for rapidly and reproducibly measuring the chemical composition of samples with little or no sample preparation*". He based his statement on the fact that each of the major chemical components of a sample has NIR absorption properties, which can be used to differentiate one component from the others. NIR diffuse reflectance signals contain information about the composition of the sample and such information can be extracted by proper mathematical treatment.

When radiation strikes a solid or granular material, part of the radiation is reflected from the sample surface (Wehling, 1994). This mirror-like reflectance is called specular reflectance, and has little useful information about the sample (Figure 1). Most of the specularly reflected radiation is directed back towards the energy source. Another portion of the radiation will penetrate through the surface of the sample and be reflected off several sample particles before it exits the sample. This is referred to as diffuse reflectance, and this diffusely reflected radiation emerges from the surface at random angles through 180°. Each time the radiation interacts with a sample particle, the chemical constituents in the sample can absorb a portion of the radiation. Therefore, the diffusely reflected

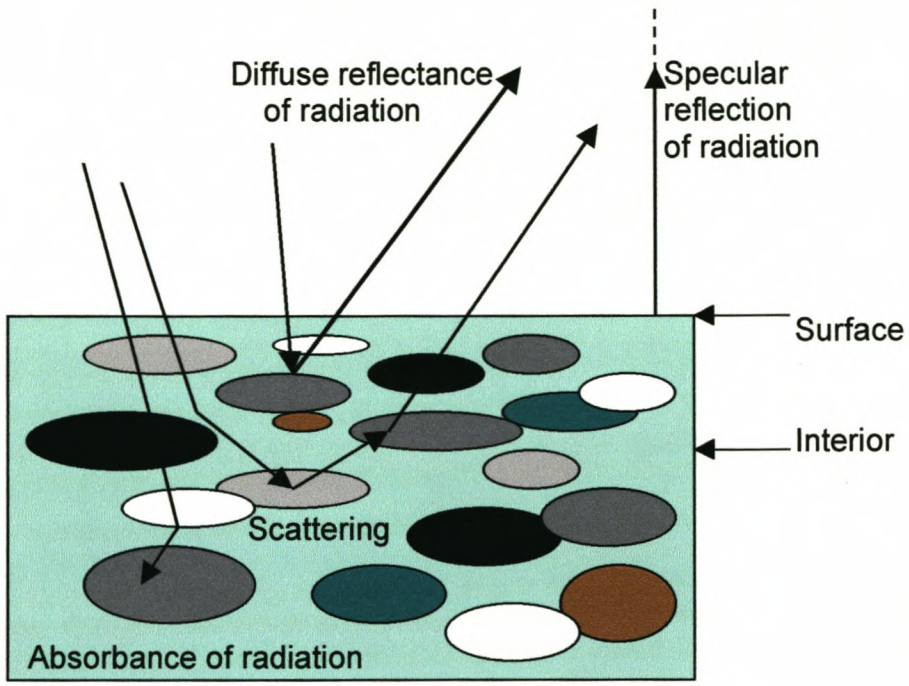


Figure 1. Diffuse reflectance of radiation (Osborne *et al.*, 1993).

radiation contains information about the chemical composition of the sample, as indicated by the amount of energy absorbed at specific wavelengths (Osborne *et al.*, 1993; Wehling, 1994).

NIR spectroscopy deals with the interaction of near infrared light with matter and a phenomena associated with light propagation, such as refraction, can be explained by using the wave theory of electromagnetic radiation called the Beer-Lambert law (Osborne *et al.*, 1993; Penner, 1994). This law states that the fraction d*P*/*P* of radiant energy *P* absorbed by an infinite thickness of sample is proportional to the number of molecules d*n* in that thickness, given by

$$-dP/P = k\, dn \qquad .....1$$

Experimentally, the fraction of radiation (P/P_0) transmitted by the sample is measured and this is called transmittance (T). In practice, transmittance is converted to the absorbance (A), which is defined by

$$A = \log 1/T = \log (P_0/P) \qquad .....2$$

$$\text{where } A = abc \qquad .....3$$

and *a* = absorptivity

b = the thickness through which the radiation passes

c = the concentration of molecules in the sample

In order to employ equation 3 to determine the concentration of a sample from its measured absorbance, it is necessary to have an accurately defined sample thickness and to determine the value of the absorptivity using a series of samples of known concentrations.

Osborne *et al.* (1993) defines the NIR wavelength range from 700 – 2500 nm (14 300 - 4000 cm^{-1}), where Wetzel (1998) states that the region lies between 850 - 2500 nm (11 000 – 4000 cm^{-1}). According to Norris (1989) the

NIR region is generally defined as the wavelengths from 700 – 3000 nm (14 285 - 3333 cm^{-1}). However he does state that most of the quantitative analyses are done in the 1200 - 2500 nm (8333 - 4000 cm^{-1}) region because absorption bands below 1200 nm are too weak and those above 2500 nm too strong, thereby complicating quantitative measurements by reflectance. Figure 2 is a diagrammatic presentation of the electromagnetic spectrum, which shows the near infrared region in relation to the other wavelength regions.

Absorption peaks arise from overtones and combination bands of fundamental vibrations in the mid-infrared spectrum from 2500 – 15 000 nm (4000 - 600 cm^{-1}) and between 3000 - 20 000 nm (3333 - 500 cm^{-1}) according to Wetzel (1998) and Williams & Stevenson (1990), respectively, resulting in absorption which tend to be weak in intensity (Wehling, 1994). This is actually an advantage, as absorption bands that have sufficient intensity to be observed in the NIR region, arise primarily from functional groups that have a hydrogen atom attached to a carbon, nitrogen or oxygen, and which are common groups in the major constituents in food. Wetzel (1998) sees the weakness of band intensity also as a virtue of NIR as he explains how the samples can be rather thick (0,1 to 1 mm) to compensate for the low intensity, which is more convenient for the researcher. This will be further explained in the section on how to interpret the spectra.

B2. Instrumentation

Filter instruments

In the early years of NIR spectroscopy the first commercial instruments introduced were based on interference filters (Stark *et al.*, 1986). The filters were utilized to select a narrow wavelength range and in that way they could measure well defined components such as protein, water or oil in certain agricultural products (e.g. wheat, and corn), as long as they had filters at a wavelength characteristic of them (Harsanyi & Varadi, 1986). Although these filter instruments are most often the best choice for routine applications for

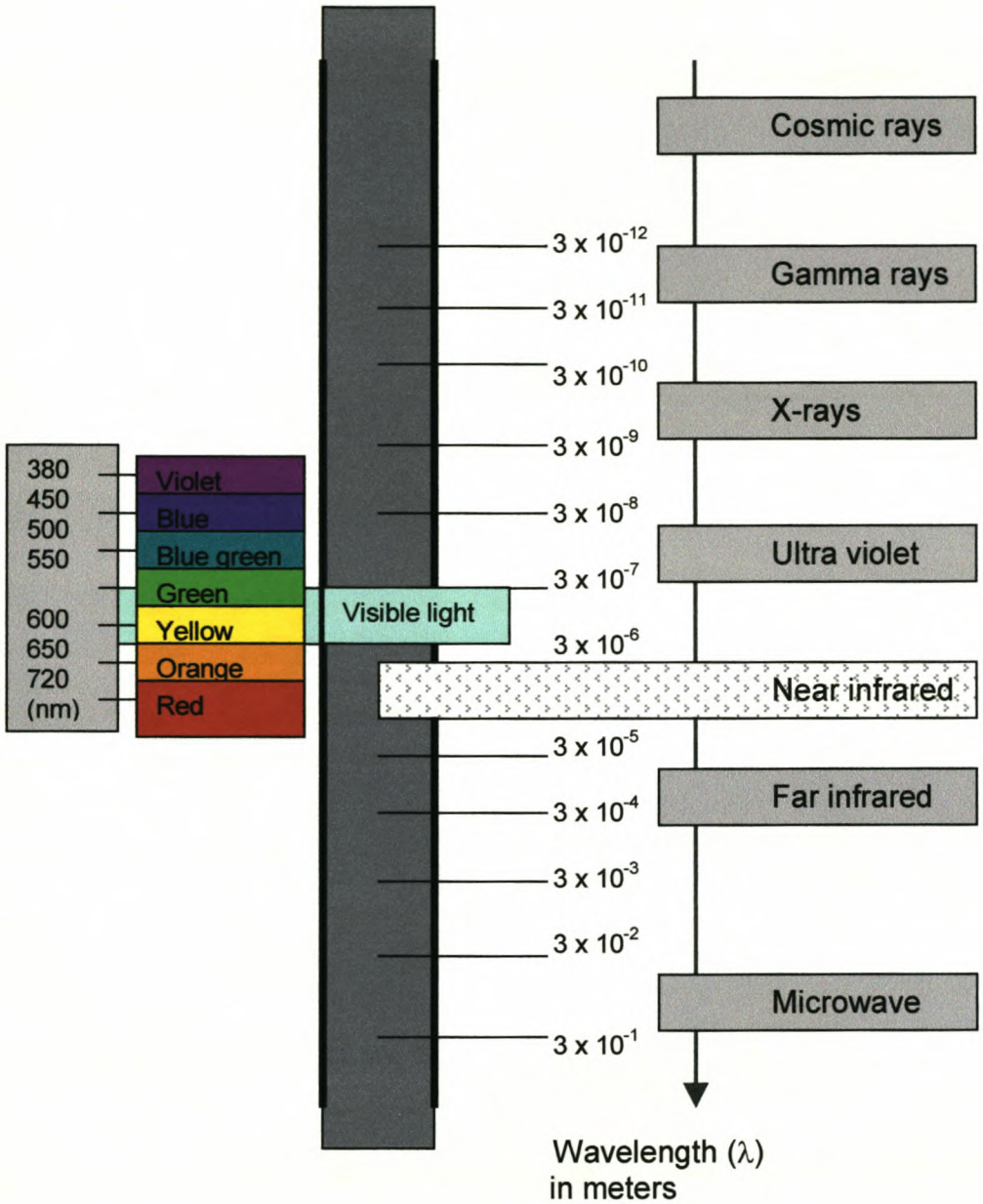


Figure 2. The electromagnetic spectrum showing the near infrared region (Penner, 1994).

reasons such as low cost, superior energy throughput and a design ideal for use outside of a laboratory setting, grating monochromators have replaced them. This is because of the lack of complete flexibility in wavelength selection desirable for some applications, particularly for research and method development. To meet this need, grating monochromators designed specifically for NIR analyses became commercially available (Stark *et al.*, 1986).

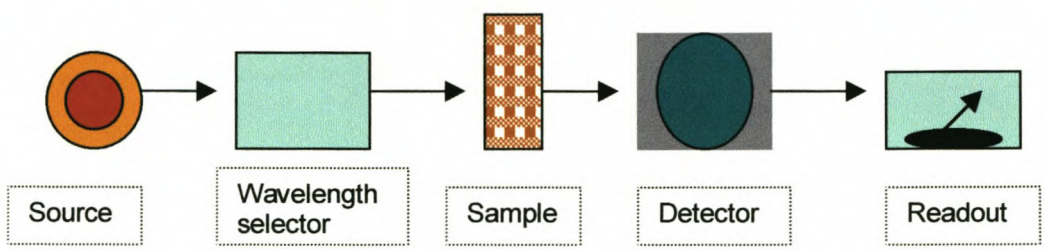
Monochromators

In the NIR region a quartz tungsten halogen filament lamp is the usual source of radiation for an interference filter instrument and it is further equipped with a PbS photoconductive cell as the detector. A phase sensitive amplifier, a measuring instrument which combines the features of sources and detectors with sample presentation, and a digital readout are also part of the basic design of such an instrument (Wehling, 1994; Wetzel, 1998; Willard *et al.*, 1988; Osborne *et al.*, 1993). Although the arrangement of these components differs between instruments, the basic configurations for transmittance and reflectance stays the same as shown in Figure 3 (Osborne *et al.*, 1993).

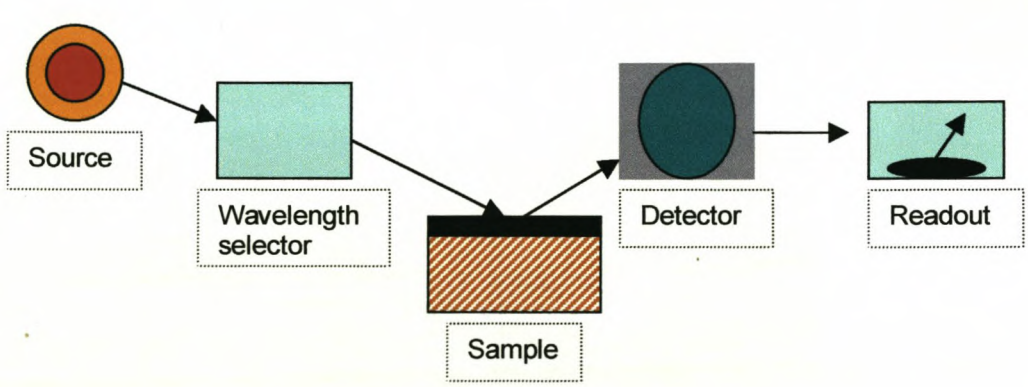
The devices described in the previous section use optical interference to produce angular dispersion. The following section deals with modulators that do not produce angular dispersion and are commonly referred to as interferometers.

Interferometers

The Michelson interferometer - Near infrared radiation can be analysed by means of a scanning Michelson interferometer. The Michelson interferometer creates the conditions for optical interference by splitting light into two beams and then recombining them after a path difference has been introduced. In such an interferometer (Figure 4), a beam of radiation from the source entering the interferometer encounters a beam splitter. This optical device reflects approximately half the incident radiation and transmits the other half. In one pathway, a second mirror is encountered, and the radiation coming off the



(a) **TRANSMITTANCE**



(b) **REFLECTANCE**

Figure 3. Basic instrument configurations for (a) transmittance and (b) reflectance (Osborne *et al.*, 1993).

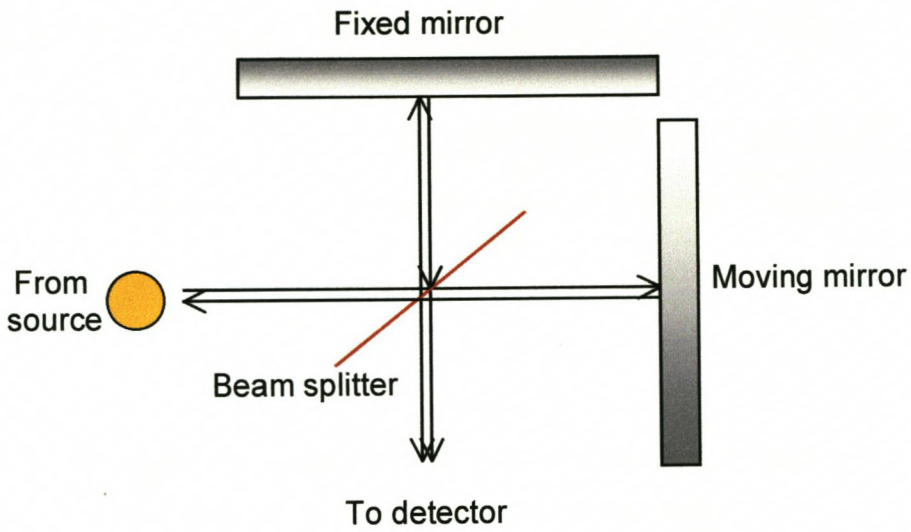


Figure 4. Michelson interferometer schematic diagram showing how the light is split into two beams and then recombined after a path difference has been introduced (Osborne *et al.*, 1993).

second mirror rejoins the rays coming straight through the beamsplitter to proceed onward to the sample target. Because the second mirror is moving, the pathway to and from the movable mirror will be variable as a function of time. At different mirror positions, a difference in the two path lengths produces interference and the data collected during the time of oscillation are subjected to fast Fourier transformation (Willard *et al.*, 1988; Osborne *et al.*, 1993; Wetzel, 1998).

The Fourier transform interferometer - Fourier transform interferometers gave life to Fourier transform spectroscopy and thus the near infrared Fourier transform (FT-NIR) spectrophotometer. Advantages of FT over dispersive spectrophotometers are: greater wavelength accuracy; greatly improved signal-to-noise ratio; increased energy throughput; and FT instruments can acquire spectra more rapidly because all wavelengths are measured at once (Wehling, 1994; Willard *et al.*, 1988). The process of Fourier transform spectroscopy is illustrated by Willard *et al.* (1988) as the signal resulting from the detector known as an interferogram (stored in memory 1), which contains all the information required to reconstruct the spectrum via the mathematical process known as Fourier transformation (Figure 5). There is an automatic process between the initiation of the scan and the final plot. This interferogram is then automatically aligned with, and added to, the averaged interferograms (stored in memory 2). At the same time annotation of the plot is begun in preparation for the final spectrum. After a number of scans, often 32 (approximately 60 seconds), this averaged interferogram is Fourier transformed to produce a single-beam spectrum (stored in memory 3). This single-beam spectrum is then rationed against the stored background (memory 4) and the resulting "double-beam" spectrum is plotted on the high-speed digital plotter. Additional space is available for storage of a reference spectrum that would be used in the spectral subtraction technique (memory 5). This additional space is also used to store a newly measured spectrum while maintaining the sample and background spectra for further manipulation. The time from insertion of the sample to the completed

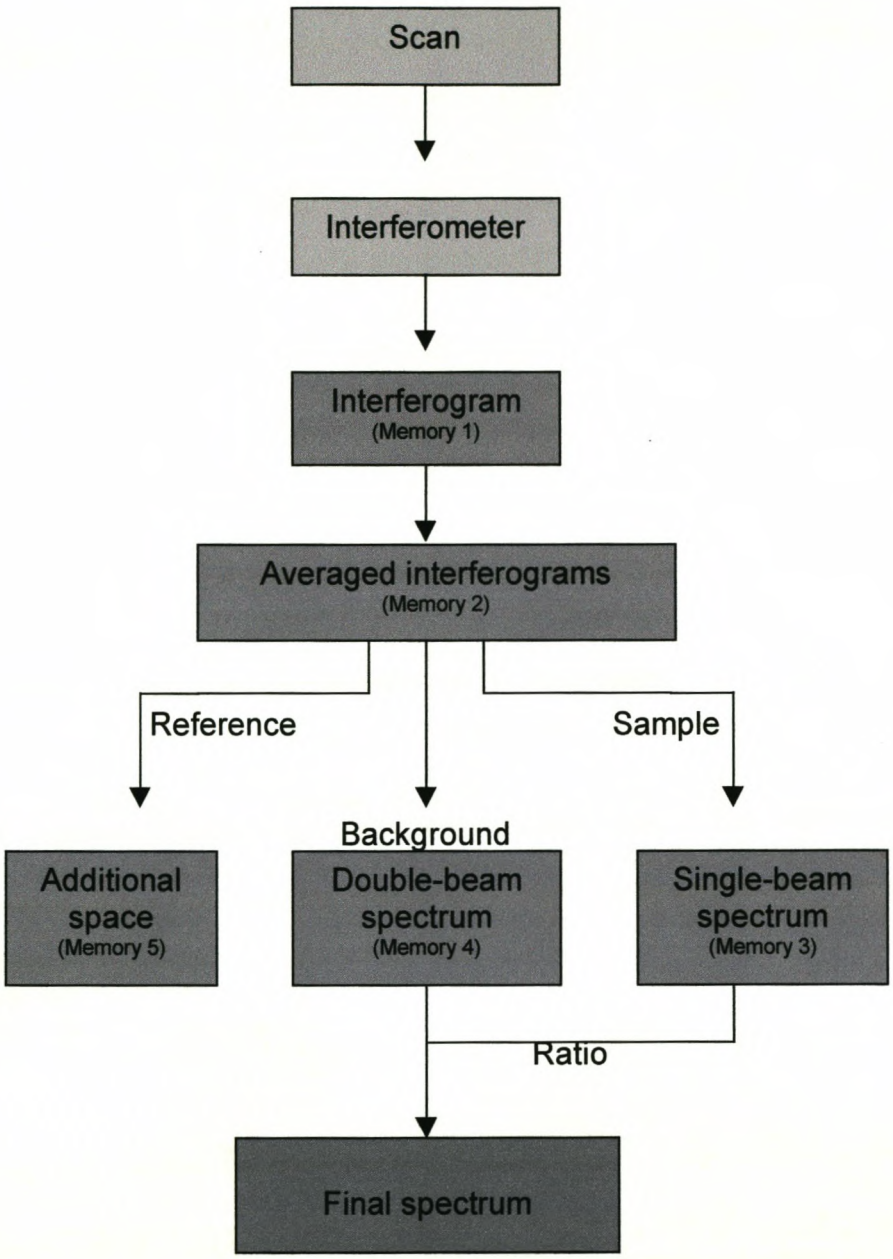


Figure 5. Block diagram of the FT-NIR spectrophotometer's functions (*Willard et al.*, 1988).

plot is about 2 minutes (Willard *et al.*, 1988). All the memories function on their own.

A decade ago FT-NIR spectrophotometers were more expensive than sequential disperse instruments due to the precision needed for mirror movement and the computers required (Willard *et al.*, 1988). Currently the cost of interferometers has decreased and because of advances in electronics the economic barrier is no longer prohibitive toward FT-NIR spectrophotometers (Wetzel, 1998).

B3. Spectral information

Although weak near infrared bands appear to be a major disadvantage, this apparent disadvantage can be a benefit to the analyst (Wetzel, 1983; Wetzel, 1998). Usually the presence of many bands makes data difficult to interpret, especially when there is considerable overlapping of the weaker, i.e., when the third overtone is weaker than the first. For example, at 2000 - 2500 nm (5000 - 4000 cm^{-1}) there are intense bands in nearly every compound containing carbon and hydrogen or nitrogen and hydrogen. At the 1000 nm (10 000 cm^{-1}) region, bands from the same vibration are barely detectable. In general, for strong bands absorbing at long wavelengths the same vibrations produce weaker bands at short wavelengths (Wehling, 1994).

In the near infrared region only bands that contain light atoms and have strong molecular bonds, are seen (Wetzel, 1983). Typical NIR spectra of wheat, dried egg white and cheese have strong absorption bands associated with the -OH groups of water which are centered at 1450 and 1940 nm (6869 and 5154 cm^{-1}) (Wehling, 1994). Bands arising from the –NH groups in protein can be observed at 2060 and 2180 nm (4854 and 4587 cm^{-1}) in egg white spectrum, but are partially obscured by a starch absorption band, centered at 2100 nm (4761 cm^{-1}), in a wheat sample. Relatively sharp absorption bands arising from –CH groups in lipid can be observed at 2310 and 2350 nm (4329 and 4255 cm^{-1}), and another band from these groups is seen around 1730 nm (5780 cm^{-1}).

Studies of the 2100 nm (4761 cm^{-1}) region indicated the presence of structures common to starch and cellulose. On either side of the band at 2100 nm (4761 cm^{-1}) are absorptions at 2055 and 2180 nm (4866 and 4587 cm^{-1}) corresponding to amide structures present in protein. At approximately 2310 and 2347 nm (4329 and 4260 cm^{-1}), prominent peaks appear for oilseed samples, indicating the presence of oil (Wetzel, 1983).

Two important features are seen in the plot of an alcoholic beverage: the absorption band 1330 - 1630 nm (7518 - 6134 cm^{-1}) which is caused by water and functional groups in sugars; and the band from 1630 - 1830 nm (6134 - 5464 cm^{-1}) contains ethanol information lifted by the tailing of the formal band (Van den Berg *et al.*, 1997). Figure 6 gives an example of a NIR spectra taken from a must samples which does not yet have any alcohol present, thus the absence of the ethanol band and an example of a wine sample with alcohol present.

B4. Sample handling

Solids

Most samples are prepared by packing the food tightly into a cell against a quartz window, thereby providing a smooth, uniform surface from which reflection can occur. Quartz does not absorb in the near infrared region. At each wavelength, the intensity of light reflecting from the sample is compared to the intensity reflected from a non-absorbing reference, such as a ceramic or fluorocarbon material or the interior of the integrating sphere (Wehling, 1994).

In the case of non-homogenous samples, it was important to thoroughly mix or homogenize the samples prior to analysis (Osborne *et al.*, 1993). In the past grains, for instance, had to be grounded before packing to make the samples as homogenous as possible, but nowadays whole grain samples can be used (Lamb & Hurburgh, 1991; Corbellini & Canevara, 1994).

Near infrared spectroscopic studies on fruit and vegetables provides the opportunity for internal quality assessment of the whole intact fruit (Kawano *et al.*,

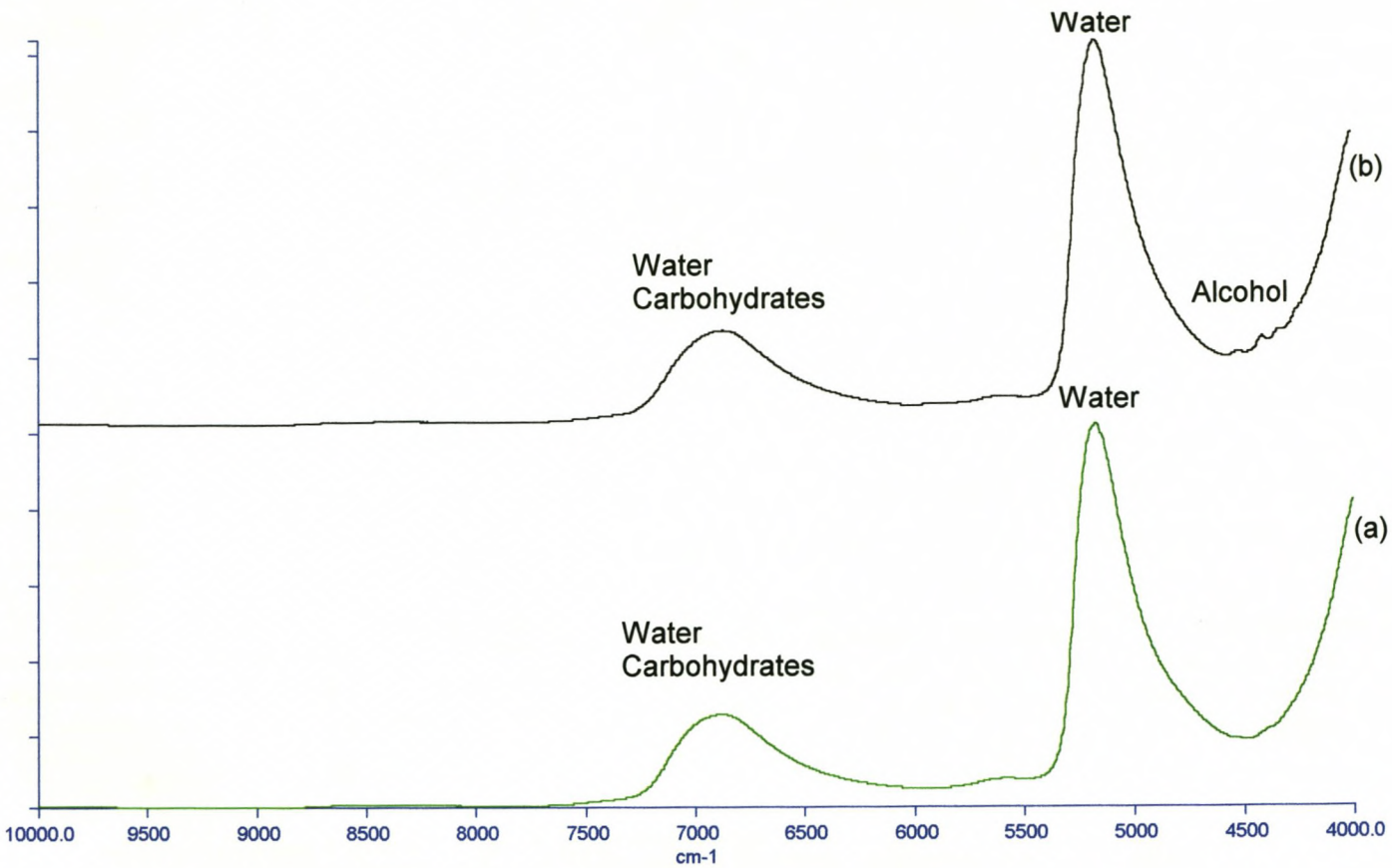


Figure 6. Near infrared spectra of a (a) must and (b) wine sample (offset for clarity).

1995; Slaughter, 1995; Peiris *et al.*, 1997). No sample preparation or homogenization of such samples is necessary, resulting in non-destructive determinations

Liquids

The measurement of liquids is generally more straightforward than the measurement of solids although wider absorption ranges may be encountered (Osborne *et al.*, 1993). The Beer-Lambert absorption law is obeyed unless the liquid is full of particular matter or extremely turbid. The width of the cuvette of flow-through cell can be adjusted to optimize the absorption sensitivity.

C. Near infrared calibration

NIR instruments, just like most other measuring instruments, require calibration before they can be used for quantitative measurement. There are a variety of calibration models in use that is implemented to overcome the same type of basic difficulties. These difficulties are due to the complex nature of NIR spectra, in which every peak of interest is surely overlapped by one or more interfering peaks. The strong dependence of reflectance on the scattering properties of the sample causes some problems. The implementation of an appropriate calibration model is thus crucial to obtain the best results (Osborne *et al.*, 1993).

Baughman *et al.* (1993) have experienced another sort of frustration concerning NIR calibration. According to them NIR calibrations are so easily executed in feasibility studies, but difficult in reality. What they refer to is the number of feasibility studies that have been published by the academic community, vendor community and users, but the problem lies with the slowness of acceptance of the method by the process control community. They came up with some explanations: In feasibility studies there is a tendency to bypass a lot of the real problems that are of major concern; and there need to be close co-operation between the vendor and the user in order to discover some of these problems.

C1. Calibration data set

For meaningful interpretation of the spectra and to avoid misleading calibration results, knowledge of the calibration process is needed (Baughman *et al.*, 1993; Osborne *et al.*, 1993; Wetzel, 1998). When conducting a calibration, the first task is to collect a set of calibration samples, irrespective of the mathematics that are to be invoked in the calibration. The calibration set should be representive of the population of samples that needs to be analysed in future with the instrument. To be successful analytically, the set must be robust, i.e., it must be applied successfully to future samples and to try to accomplish this, any variables anticipated in the future should be incorporated into the database.

C2. Data reduction and pretreatment

Spectral data must be transformed to reduce the large number available from the monochromator's data but without loss of information (Osborne *et al.*, 1993; Wetzel, 1998). Among these transforms are principal component regression (PCR) and partial least squares (PLS), known as global methods. The essence of these methods lies in their construction of factors from the original spectral data and involve no wavelength selection (Osborne *et al.*, 1993, Wetzel, 1998). Partial least squares, however, do utilize spectral as well as reference data in developing calibrations.

Apart from the above-mentioned transformation, the use of derivatives (differences) is useful because it substantially reduces, although not entirely removes, particle size effects (Osborne *et al.*, 1993). It is a useful approach of solving two problems concerning near infrared spectra, namely overlapping peaks and baseline shifts (Buco, 1994). The second derivative of a spectrum is able to separate overlapping absorption bands and will also remove baseline shifts. The principle is based on the fact that the derivative of a straight line is zero and therefore the derivative of a spectrum-plus-a-straight-baseline will be the same as the derivative of the spectrum. The first derivative have the same

effects, but to a lesser extent. Higher than the second derivative tend to be noise sensitive and generate more artifacts than the lower order derivatives.

Another spectral pretreatment technique is that of multiple scatter correction (MSC) which was developed to remove unwanted variability due to particle size or path length variation (Buco, 1994; Fearn, 1999a). MSC rotates the spectra to remove some of the effect of particle size on scattering (Osborne *et al.*, 1993). This technique rotates each spectrum so that it fits as closely as possible to the mean spectrum. The MSC theoretically ignores chemical effects on the light scatter (Buco, 1994).

Multiple linear regression (MLR)

Multiple regression by least squares is a simple generalization of a straight line fit (Osborne *et al.*, 1993). When performing multiple linear regression (MLR), Fourier coefficients rather than log 1/R spectral data can be used. This method is preferred because of the speed of computation of data and the reduction of data storage requirements. The use of Fourier coefficients also avoids the problems of wavelength selection and intercorrelation present with log 1/R data, and more terms can be used in the equation without overfitting (Martin, 1992; Wetzel, 1998). A particular problem with MLR using log 1/R data is caused by the high correlations between log 1/R at different wavelengths (Osborne *et al.*, 1993).

Principal component regression (PCR) and partial least squares (PLS)

Principal component regression and PLS start with a principal component transform in which the variance in the data is explained. The process involves orthogonal axes and the weights are expressed in eigenvectors. The first principal component accounts for more than 90% of the variance. Each successive principal component accounts for less than the previous one (Wetzel, 1998). The difference between PCR and PLS lies in the manner in which the spectra are decomposed into factors. In PCR, decomposition is performed

independently of constituent concentration data. In PLS, the concentration information is related to the factoring (Martin, 1992).

According to Osborne *et al.* (1993) the essence of PCR and PLS lies in their construction of factors from the original spectral data, keeping in mind the aim of reduction of spectral data without discarding useful information and by doing so avoiding overfitting.

Principal component regression is a form of data compression where the data are decomposed into new variables that are linear combinations of the original data (Martin, 1992). The manner in which the new variables (principal components or factors) are created can be visualized for a two-dimensional system. If the log (1/R) data of a set of samples at two wavelengths are plotted against each other, a new axis (principal component) can be found in the direction of maximum variability of the data. If more wavelengths are added, the plot becomes multidimensional and more axes can be found, each one successively accounting for the maximum possible amount of the remaining variability, and each orthogonal to the other. The total variance accounted for in each factor is its eigenvalue (Martin, 1992).

Wetzel (1998) explains how PCR provides useful information by stating that if the magnitude of the eigenvector of the first principal component is constant for all wavelengths, it is likely that there is a non-chemical cause for the variance described by that principle component. Martin (1992) finds the fact that all the components are orthogonal to each other, and consequently eliminating collinearity without eliminating spectral information, as a major advantage because it makes the models more robust. Noise is generally reduced because it is spread throughout all the factors, while the variation of interest is concentrated into the first few factors.

The computations of PLS are usually performed intuitively, calculating the variables in the order of importance (only the number of latent variables that are useful in the model are retained), with factors being constructed and added to the regression one at a time until the procedure is stopped (Osborne *et al.*, 1993). It is important to know when to stop adding the components, because overfitting

(calibrations which are too calibration-data dependent) must be avoided. Performance of the calibration model will continue to improve for as long as factors are added, but models with many factors rarely predict well. Using too many components results in overfitting and too few in underfitting, both leading to poor predictions. Protection against these phenomena is provided by using three sample sets: either by the use of an optimisation set (factors being added until performance on the validation set is optimised); a validation set (determine predictive accuracy of a single, selected model); or more commonly, by partial or full-cross validation (Wetzel, 1998; Osborne *et al.*, 1993; Martin, 1992; Næs & Isaksson, 1991a). When using PLS it is also important to remember that least squares are very sensitive to outliers. If a residual is large, its square is very large, and can easily dominate the sum of squares. It is thus very important to check for, and if necessary, remove outliers (Fearn, 1999b).

C3. Validation

Once a NIR method has been developed by calibrating against a reference method, it can be used to measure some chemical constituents or physical properties of a different set of samples know as the validation set (Buco, 1994). The comparison of the measurements of the constituents or properties for the NIR method with the measurements from the reference method is called validation of the calibration process or prediction testing (Buco, 1994; Næs & Isaksson, 1991b).

Prediction testing is based on splitting the data set in two, one for calibration and the other for validation (Næs & Isaksson, 1991b). Cross validation is a technique based on the calibration data only. It is similar to prediction testing since it only tests predictors on data that were not involved in calibration, but for full-cross validation this is done by successively deleting samples from the calibration set. First, sample number one in the calibration set is deleted and then the calibration is performed on the rest of the samples before it is tested on the first sample by comparing y with $\hat{y}$. The first sample is then put back into the calibration set, and the procedure is repeated by deleting sample

number two. The procedure continues until all samples have been deleted once (Næs & Isaksson, 1991b).

C4. Statistical correlations

Several statistics can be generated to assess the goodness of a calibration model, among them the correlation coefficient (r), the standard error of calibration (SEC), the F-statistic, the root mean square error of prediction (RMSEP) and the root mean square error of cross validation (RMSECV).

The degree of linear relationship (or the degree of association) between two variables is often described by the correlation coefficient, r (Buco, 1994). In the case of a calibration model the two variables present are y, the reference value, and $\hat{y}$, the predicted value which are incorporated in the equation for the correlation coefficient (equation 4).

When considering the criteria of successful regressions, a correlation coefficient in the vicinity of 0.9, but preferably approaching 0.999, is good (Wetzel, 1998).

$$r = \frac{\sum_{i=1}^{n_s}(\hat{y}_i - y_i)^2}{\sqrt{\sum_{i=1}^{n_s}(y_i - \bar{y})^2}} \qquad \text{.....4}$$

The standard error of calibration (SEC) or the F-statistic is perhaps a better indicator of the success of the goodness of fit of the line. SEC is a root mean square average of the errors about the fitted line and as such represents a typical discrepancy from the line (Osborne *et al.*, 1993). In other words, SEC (equation 5) indicates the standard error of difference between calculating the analyte predicted value from the empirical equation produced and the analyte value from some reference method, thus how well the calibration samples were fit (Wetzel, 1998; Westerhaus, 1989). The overall performance of the model is evaluated by the F-statistic (equation 6) evaluate which is calculated to

determine whether the property variance significantly accounts for by the model or not.

$$SEC = \sqrt{\frac{1}{n-1-t}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad .....5$$

With: y_i = the reference value for the i^{th} sample

$\hat{y}_i$ = the predicted value for the i^{th} sample

n = the number of samples

t = the number of terms in the model

$$F = \frac{\sum_{i=1}^{n_s}(y_i - \hat{y})^2(n_s - n_f - 1)}{\sum_{i=1}^{n_s}(y_i - \hat{y}_i)^2(n_f - 1)} \qquad .....6$$

where n_s is the number of significant factors in the model. This is the F-value found in the standard analysis of variance (ANOVA) table model. A poor regression will give a low (<3) value for F. The F-value can be viewed as a measure of the signal to noise in the model (Anonymous, 1997).

Prediction testing is based on splitting the data set into two, one for calibration and the other for validation (Næs & Isaksson, 1991). The prediction testing estimate is called the root mean squared error of prediction, RMSEP, and can be written as equation 7 (Næs & Isaksson, 1991).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad .....7$$

With: y_i = the reference value for the i^{th} sample

$\hat{y}_i$ = the predicted value for the i^{th} sample

n = the number of samples

RMSEP is an estimate of the accuracy of the calibration against the reference method, and is calculated using an independent test set. An important advantage with prediction testing is that it is the only method available that estimates the prediction ability of the actual predictor determined by the actual calibration data (Næs & Isaksson, 1991c). Alternatively, the accuracy of a calibration can be expressed as standard error of prediction, SEP (equation 8) of the bias-corrected residuals (equation 9). The SEP value gives an estimate of the magnitude of the error expected when independent samples are predicted using the model. In effect, a standard is removed from the multiple linear regression and a model built using the other n_s - 1 standards. The removed standard is predicted using this model. This is done for each standard in the calibration set (Næs & Isaksson, 1991c; Osborne *et al.*, 1993; Delwiche, 1993).

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i - BIAS)^2}{n-1}} \qquad .....8$$

Where:

$$BIAS = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \qquad .....9$$

The BIAS is interpreted as the average difference between y and $\hat{y}$ in the prediction set (Anonymous, 1997). If the BIAS is near a value of zero, the overall error of validation (SEP) can be interpreted as the standard deviation (SD) of the NIR prediction.

When cross validation is performed the accuracy can be calculated as a root mean square error of cross validation (RMSECV) as illustrated in equation 10 (Næs & Isaksson, 1991c).

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad \text{.....10}$$

With: y_i = the reference value for the i^{th} sample

$\hat{y}_i$ = the predicted value for the i^{th} sample when it was dropped from the ML regression

n = the number of samples

Alternatively the accuracy of a calibration can be expressed as standard error of cross validation, SECV (equation 11) of the bias-corrected residuals (Anonymous, 1997).

$$\text{SECV} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i - BIAS)^2}{n-1}} \qquad \text{.....11}$$

The standard deviation of the reference data (SD) divided by the SEP or SECV is called the standard deviation of reference data (RPD) (Williams, 1991). This is an indication of the efficiency of a calibration and defined by equation 12. The applications of the RPD values are explained in Table 1.

$$\text{RPD} = \frac{SD}{SEP} \text{ or } \text{RPD} = \frac{SD}{SECV} \qquad \text{.....12}$$

Table 1. Interpretation of RPD statistics (P.C. Williams, Canadian Grain Commission, personal communication).

RPD value	Classification	Application
0.0 - 2.3	Not recommended	-
2.4 - 3.0	Poor	Very rough screening
3.1 - 4.9	Fair	Screening
5.0 - 6.4	Good	Quality control
6.5 - 8.0	Very good	Process control
8.1+	Excellent	Any application

C5. Soft independent modeling by class analogy (SIMCA)

NIR spectroscopy can be used for classification and verification of raw materials. If the materials to be identified are spectroscopically dissimilar, it is often only necessary to use a simple distance measure such as a spectral difference (Anonymous, 1998). If the spectra are similar, it may be necessary to include slightly more sophisticated techniques that take into consideration both the variability of the spectra of interest and the differences between the spectra. The SIMCA (Soft independent modeling by class analogy) technique provides such a method.

In the SIMCA method, a principal component model is created for each class (Anonymous, 1997). An envelope to contain the standards of a class is constructed. If one principal component is used, the class mean and a line can model the data. If two principal components are used, the class mean and a plane are required, and with three principal components, the class mean and a volume are required. If there are more than three components, the envelope can be thought of as a box with more than three dimensions (a hyperbox). In a SIMCA classification the unknown spectrum is classified by whether it lies inside or outside the hyperbox (this is the model residual).

An advantage of SIMCA is the use of an objective statistical test, the F-test, to establish the probability of a sample belonging to any given class. SIMCA is thus

a method that provides a set of parameters that characterise each class and are the basis for other quantities that describe the data. The procedure checks every standard spectrum to ensure that the ones from a single class fit that class (recognition), and that those from other classes selected are rejected (rejection) (Anonymous, 1998).

D. Applications of near infrared analysis in the food and beverage industries

Very few substances or materials can not be analysed by means of near infrared spectroscopy. All applications to be mentioned prove the point that NIR spectroscopy has a broad application field in the food and beverage industries. According to Williams & Stevenson (1990) from the Canadian Grain Commission's Grain Research Laboratory, the applications of NIR in the food industry were the following up to the beginning of the 90's:

1) analysis of raw materials to ensure that raw materials are capable of yielding the required amounts of products, and that specifications can be met in the finished product for parameters such as protein and fiber contents;
2) analysis of the product at critical stages during processing to monitor composition;
3) analysis of the finished product to verify that specifications have been met;
4) analysis of by-products to determine their value for future sale; and
5) analysis of effluents and other wastes for pollution control.

Although the above mentioned applications are still the major ones of NIR spectroscopy in the food industry, the source materials on which the principles are applied, have multiplied in recent years as a result of intensive research. The application of NIR in the food industry originated with agricultural research when NIR spectroscopy was applied to a wide selection of seeds to determine the moisture thereof (Ben-Gera & Norris, 1968). Nowadays NIR has found a niche in many (almost every) area of the food and beverage industries as summarized by the following examples of applications of recent years.

Flour industry

Wheat kernels, the elementary components of flour, is the object of NIR spectroscopic research for prediction of the presence of insect larvae in individual wheat kernels (Ridgway & Chambers, 1996; Ghaedian & Wehling, 1997). In both studies it was found that the NIR method could differentiate between uninfected samples and samples with insects present, concluding thus that NIR should be useful as a rapid method of detection. Uthayakumaran *et al.* (1998) determined the NIR spectral differences between gliadin and glutenin proteins, thereby indicating the potential of estimating these properties in whole grain.

Different commercial flour types from a number of mills were collected to develop a discriminate model for a quantitative analysis, resulting in a classification rate of 97% (Sirieix & Downey, 1997). This application of NIR analysis is thus very useful as a wheat flour authentication method. NIR spectroscopy is also applied as a physical process to assure high-quality flour after milling. This post-milling, pre-storage treatment is necessary to ensure a pure product, as consistently as possible, and NIR has found its niche as an on-line scanning system to monitor and control the performance of the mill (Sugden, 1997).

Dough development is another study field concerning the flour industry and Wesley *et al.* (1998b) did intensive research on the monitoring of dough mixing and development by NIR spectroscopy. NIR spectroscopy showed potential to provide information on the chemical processes that occur during dough development in relation to flour strength, mixing action and intensity.

Rice industry

Kawamura *et al.* (1997a) determined the constituent content of undried Japanese rice by use of NIR transmission spectroscopy. Near infrared technology were adequate to classify the undried rice into qualitative groups (high protein and low protein content rice) on reception at the rice drying facility. This study also included the determination of physiochemical properties of rice and the results

showed that the NIR technology is successful in assessing rice quality (Kawamura *et al.*, 1998). Another independent study focussed on the replacement of sensory tests with a calibration model for the evaluation of rice taste (Kawamura *et al.,* 1997b). However, the results was not adequate to justify the replacement of sensory tests with a NIR technique, although NIR technology could be used to classify the rice samples into qualitative groups, such as poor taste, better taste and the best taste. The feasibility of using NIR spectroscopy to discriminate between Basmati and other long-grain samples was investigated by Osborne *et al.* (1997) and reasonable results were obtained.

Coffee industry

A number of studies have been done on coffee beans (Downey *et al.*, 1994; Downey & Spengler, 1996) analysing the components and authenticating whole and ground coffee beans to of identifying coffee variations using NIR spectroscopic techniques. They concluded that the study proved potentially useful as an accurate classification method of whole coffee beans and ground coffee beans (Downey *et al.*, 1997). The study by Downey *et al.* (1997) revealed that NIR spectroscopies have the potential to discriminate between different coffees, for example between Arabica and Robusta.

Fats and oils

The nondestructive determination of fatty acid composition of a variety of seeds (soybeans, sunflower, *Brassica napus, Brassicaceae*) and other selected vegetables have been studied by means of NIR spectroscopy (Sato *et al.*, 1994; Sato *et al.*, 1995; Sato, 1997; Pazdernik *et al.*, 1997; Velasco *et al.*, 1998; Sato *et al.*, 1998; Velasco *et al.*, 1999). Principal component analysis on NIR spectroscopic data for classification purposes of vegetable oils and other essential oil plants were applied (Sato, 1994; Schulz *et al.*, 1998). NIR methods showed the feasibility to measure quality criteria for the purpose of process

control in the plant instead of the time-consuming chemical analyses (Sato *et al.*, 1994; Velasco *et al.*, 1998, 1999).

Meat industry

NIR spectroscopy is applied in the meat industry mainly as a discrimination method and for authentication purposes. Differences between fresh and frozen beef can be detected by NIR spectroscopy techniques (Downey & Beauchêne, 1996, 1997) and discrimination between raw pork, chicken and turkey meat has been studied (Rannou & Downey, 1997). The possibility to determine the sodium chloride content of sausages by NIR spectroscopy was studied by Ellekjær *et al.* (1997) showing promising results.

Dairy industry

A great number of studies have been done on NIR spectroscopic techniques applied in the dairy industry's different sectors: the powder dairy industry; liquid milk; cheese; butter; and fermented milk products as summarised in a review by Rodriguez-Otero *et al.* (1997a). NIR spectroscopy has been applied as an adulteration detection technique that looks for the presence of strange fats in milk samples (Sato *et al.*, 1990). NIR spectroscopy has also been applied to detect substances that can be added to the milk such as water, sodium chloride and skim milk powder (Pedretti *et al.*, 1993). Rodriguez-Otero *et al.* (1997b) used NIR transflectance spectroscopy with success to determine fat, protein and total solids of fermented milks avoiding any pretreatment of the samples.

Wine and distillation industries

The focus of NIR spectroscopy in the distillation industry is on the use thereof as a quality verifying method (Gomez-Cordoves & Bartolome, 1993). Phenolic composition of distillates aged in wood has been studied and thus giving rise to different standards that establish their genuiness and quality. By use of NIR methods, quality classification of brandies and cognacs are obtained by use of

principal component analysis, which is applied to the results of simple chemical (pH, phenol content) and physical (colour) determinations. NIR spectroscopy is also applied to simultaneously determined routine analysis that includes ethanol, methanol, glycerol, fructose, glucose and total residual sugars present during the production of alcoholic beverages. The most common application of NIR spectroscopy in the wine industry has been for the determination of the alcohol content in wine (Gishen & Dambergs, 1998). Whilst NIR spectroscopy has been successfully used for the determination of sugars in various matrices, wine has not been one of them. The application of NIR spectroscopy for the determination of residual sugar (0 - 10 $g.L^{-1}$) in finished wine has been limited due to the lack of success in obtaining reliable calibrations (Gishen & Dambergs, 1998). Research done by Van den Berg *et al.* (1997) focussed on ethanol determination during the production of alcoholic beverages and Garcia-Jares & Medina (1996) applied NIR techniques to the routine determination of botrytized-grape sweet wines.

These evaluations of the application of NIR spectroscopy for the measurement of quality indicators in spirits and wines and the classification and grading thereof shows considerable promise to these industries. It has the potential to considerably reduce analytical times of the measurements widely used in the determining of the composition of wines and spirits.

Fruits and vegetables

NIR spectroscopy has found many applications in the fruits and vegetables industries, mainly because the technique is a nondestructive method of analysis, but also because NIR spectroscopy can be applied as an authentication technique (Kays, 1999).

In potatoes NIR has been applied to assess the total nitrogen in the potato tissue (MacKerron *et al.,* 1997) and to perform analysis of dry matter and sugars in the whole tubers (Scanlon *et al.*, 1997). The internal quality of apples have been studied by NIR spectroscopy in terms of sugar content (Moons *et al.*, 1997) and six quality factors (firmness, refractive index, pH, titratable acid, dry matter and alcohol insoluble solids content) (Zlobsdz *et al.*, 1997). Research on non-

invasive assessment of pineapple and mango fruit quality was done by Guthrie & Walsh (1997) and on intact melon and pineapples by Guthrie & Wedding (1998), using NIR spectroscopy.

Kawano *et al.* (1992) determined the sugar content in intact peaches and satsuma mandarins (Kawano *et al.*, 1993) using fiber optics as a NIR spectroscopic technique. A calibration equation was later developed with temperature compensation to determine the °Brix value in intact peaches (Kawano *et al.*, 1995). Peiris *et al.* (1997) used NIR spectroscopy to determine the soluble solids content of peaches and concluded that feasible applications of the method include packing house sorting of fruit for sweetness and parent and progeny quality assessment in peach breading programs. Various research have also been done on prediction of firmness, dry-matter, soluble solid and eating time of kiwifruits (Jordan *et al.*, 1997; McGlone *et al.*, 1997; McGlone, 1998) using NIR spectroscopy. Harvest time prediction of eating time properties of kiwifruit (Jordan *et al.,* 1997) and pre-harvest maturity by use of NIR spectroscopy of fresh dates (Schmilovitch *et al.,* 1997) was studied. A major problem for date growers is how to decide on a harvesting schedule of a plantation of fresh dates. The determination of the maturity of fresh dates and the development of a combined NIR spectrophotometer with a step-wise cell conveyor, both controlled by a personal computer, were studied by Schmilovitch *et al.* (1999). The apparatus was tested and is currently in operation, testing 100 dates in 3 minutes.

The potential for NIR spectroscopic grading of raisin quality and moisture has been studied by Huxsoll *et al.* (1995). High positive correlation was found between measured values and those predicted by NIR transmittance for bulk density, visual grade and moisture content. The results showed a relationship between raisin quality and some chemical constituents, which gives NIR transmittance the potential for measuring general raisin quality as well as specific constituents.

The technology of diffuse reflectance, illustrated in the form of near infrared spectroscopy, finds many useful applications in everyday life, but goes about it in

an unnoticed way. Many industries benefit from and use NIR spectroscopy as it offers a method for non-destructive evaluation. Such a non-destructive method allows the measurement of internal quality parameters and sophisticated chemometric tools make classification and discrimination between samples possible. The application possibilities are multiple and illustrated with success in the food and beverage industries.

E. References

Anonymous. (1997). *Spectrum QUANT+ User's Reference.* The Perkin Elmer Corporation. Beasconsfield, United Kingdom: Technical Publication.

Anonymous. (1998). A comparison of fiber optic and ICRA sampling for soft independent modeling by class analogy (SIMCA) classification of different celluloses. The Perkin Elmer Corporation. Beaconsfield, United Kingdom: FT-NIR Spectroscopy application notes.

Baughman, E.H., Van Hare, D.R. & Vickers, G.H. (1993). NIR calibration: Why is it so hard in reality and so easy in feasibility studies? *NIR news*, **4**, 10-11.

Ben-Gera, I. & Norris, K.H. (1968). Direct spectrophotometric determination of fat and moisture in meat products. *Journal of Food Science*, **33**, 64-67.

Buco, S.M. (1994). Preconference course in multivare statistics in diffuse reflectance spectroscopy. Baton Rouge, USA.

Corbellini, M. & Canevara, M.G. (1994). Estimate of moisture and protein content in whole grains of bread wheat (*T. Aestivum L.*) by near infrared reflectance spectroscopy. *Italian Journal of Food Science*, **1**, 95-102.

Delwiche, S.R. (1993). Measurement of single-kernel wheat hardness using near-infrared transmittance. *Transactions of the ASAE*, **36**, 1431-1437.

Downey, G., Boussion, J. & Beauchêne, D. (1994). Authentication of whole and ground coffee beans by near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy*, **2**, 85-92.

Downey, G. & Beauchêne, D. (1996). Authentication of fresh vs. frozen-then-thawed beef by near infrared reflectance spectroscopy of dried drip juice. *Lebensmittel-Wissenschaft &-Technologie*, **30**, 21-726.

Downey, G. & Spengler, B. (1996). Compositional analysis of coffee blends by near infrared spectroscopy. *Irish Journal of Agricultural and Food Research*, **35**, 179-188.

Downey, G. & Beauchêne, D. (1997). Discrimination between fresh and frozen-then-thawed beef *m. longissimus dorsi* by combined visible-near infrared reflectance spectroscopy: a feasibility study. *Meat Science*, **45**, 353-363.

Downey, G., Briandet, R., Wilson, R.H. & Kemsley, K. (1997). Near- and mid infrared spectroscopies in food authentication: coffee varietal identification. *Journal of Agricultural Food Chemistry*, **45**, 4357-4360.

Ellekjær, M.R., Hildrum, K.I., Næs, T. & Isaksson, T. (1997). Determination of the sodium chloride content of sausages by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **1**, 65-77.

Fearn, T. (1999a). A look at some standard pre-treatments for spectra. *NIRnews*, **10**, 10-11.

Fearn, T. (1999b). Least squares. *NIR news*, **10**, 7,13.

Foulke, F.E. (1993). Urethane in alcoholic beverages under investigation. *FDA Consumer*. http://vm.cfsan.fda.gov/~frf/fc0293ur.html. August 1999.

Garcia-Jares, C.M. & Medina, B. (1996). Application of multivariate calibration to the simultaneous routine determination of ethanol, glycerol, fructose, glucose and residual sugars in botrytized-grape sweet wines by means of near-infrared reflectance spectroscopy. *Fresenius Journal of Analytical Chemistry*, **357**, 86-91.

Ghaedian, A.R. & Wehling, R.L. (1997). Discrimination of sound and Granary-Weevil-larva-infested wheat kernels by near-infrared diffuse reflectance spectroscopy. *Journal of AOAC International*, **80**, 997-1005.

Ghishen, M. & Dambergs, B. (1998). Some preliminary trials in the application of scanning near infrared spectroscopy (NIRS) for determining the

compositional quality of grape, wine and spirits. *The Australian Grapegrower & Winemaker*, Annual Technical Issue 1998, 43-47.

Gomez-Cordoves, C. & Bartolome, B. (1993). Application of principal component analysis to simple determinations of brandies as a means of verifying quality. *Zeitschrift fur Lebensmittel Untersuchung und Forschung*, **197**, 260-263.

Guthrie, J. & Walsh, K. (1997). Non-invasive assessment of pineapple and mango fruit quality using near infra-red spectroscopy. *Australian Journal of Experimental Agriculture*, **37**, 253-263.

Guthrie, J. & Wedding, B. (1998). Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy*, **6**, 259-265.

Harsanyi, G. & Varadi, M. (1986). Technical performance of NIR instruments. International NIR/NIT Conference. Budapest, Hungary.

Huxsoll, C.C., Bolin, H.R. & Mackey, B.E. (1995). Near infrared analysis potential for grading raisin quality and moisture. *Journal of Food Science*, **60**, 176-180.

Jordan, R.B., Kunnemeyer, R., Osborne, S.D. & Seelye, R.J. (1997). Harvest time prediction of eating time properties of kiwifruit using NIR transmission. In: Proceedings from the Sensors for Nondestructive Testing International Conference and Tour, Pp. 101-110. Orlando, Florida.

Kays, S.J. (1999). Non-destructive quality evaluation of intact, high-moisture products. *NIRnews*, **10**, 12-17.

Kawamura, S., Natsuga, M. & Itoh, K. (1997a). Determining undried rough rice constituent content using near-infrared transmission spectroscopy. ASAE Annual International Meeting, Minneapolis, Minnesota.

Kawamura, S., Natsuga, M. & Itoh, K. (1997b). Visual and near-infrared reflectance spectroscopy for rice taste evaluation. *American Society of Agricultural Engineers*, **40**, 1755-1759.

Kawamura, S., Natsuga, M. & Itoh, K. (1998). Visual and near-infrared reflectance spectroscopy for determining physicochemical properties of rice. ASAE Annual International Meeting, Orlando, Florida.

Kawano, S., Watanabe, H. & Iwamoto, M. (1992). Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in interactance mode. *Journal of Japanese Society of Horticultural Science*, **61**, 445-451.

Kawano, S., Fujiwara, T. & Iwamoto, M. (1993). Nondestructive determination of sugar content in satsuma mandarin using near infrared (NIR) transmittance. *Journal of Japanese Society of Horticultural Science*, **62**, 465-470.

Kawano, S., Abe, H. & Iwamoto, M. (1995). *Journal of Near Infrared Spectroscopy*, **3**, 211-218.

Lamb, D.T. & Hurburgh Jr., C.R. (1991). Moisture determination in single soybean seeds by near-infrared transmittance. *American Society of Agricultural Engineers,* **34**, 2123-2128.

MacKerron, D.K.L., Young, M.W. & Davies, H.V. (1997). Factors influencing the calibration of near infrared reflectometry applied to the assessment of total nitrogen in potato tissues. II. Operator, moisture and maturity class. *Journal of Near Infrared Spectroscopy*, **3**, 167-174.

Martin, K.A. (1992). Recent advances in near-infrared reflectance spectroscopy. *Applied Spectroscopy Reviews*, **27**, 337-343.

McGlone, V.A., Abe, H. & Kawano, S. (1997). Kiwifruit firmness by near infrared light scattering. *Journal of Near Infrared Spectroscopy*, **5**, 83-89.

McGlone, V.A. & Kawano, S. (1998). Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology*, **13**, 131-141.

Moons, E., Dubois, A., Dardenne, P. & Sindic, M. (1997). Nondestructive visible and NIR spectroscopy for the determination of internal quality in apple. In: Proceedings from the Sensors for Nondestructive Testing International Conference and Tour, Pp. 122-132. Orlando, Florida.

Murray, I. (1999). NIR spectroscopy of food: simple things, subtle things and spectra. *NIR news*, **10**, 10-12.

Næs, T. & Isaksson, T. (1991a). Multicollinearity and the need for data reduction. *NIR news*, **2**, 10-11.

Næs, T. & Isaksson, T. (1991b). Fitting, prediction testing, cross-validation or leverage correction. *NIRnews*, **2**, 10-11.

Næs, T. & Isaksson, T. (1991c). SEP or RMSEP, which is best? *NIR news*, **2**, 16.

Norris, K.H. (1989). Definition of NIRS Analysis. In: *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality* (edited by G.C. Marten, J.S. Shenk & F.E. Barton II). Pp. 6, 12. Agriculture Handbook No. 643. United Stated Department of Agriculture: Agriculture Research Service.

Norris, K.H. (1999). Moving NIR into the next century. *NIRnews*, **10**, 4-6.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with practical applications in food and beverage analysis*, 2nd ed. Pp. 1-220. England: Longman Group UK Limited.

Osborne, B.G., Mertens, B., Thompson, M. & Fearn, T. (1997). The authentication of Basmati rice using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **1**, 77.

Pazdernik, D.L., Killam, A.S. & Orf, J.H. (1997). Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. *Agronomy Journal*, **89**, 679-685.

Pedretti, N., Bertrand, D., Semenou, M,Robert, P. & Giangiacomo, R. (1993). Applications of an experimental design to the detection of foreign substances in milk. *Journal of Near Infrared Spectroscopy*, **1**, 174-184.

Penner, M.H. (1994). Basic principles of Spectroscopy. In: *Introduction to the chemical analysis of foods* (edited by S.S. Nielsen). Pp. 317-324. Boston, MA: Jones and Bartlett Publishers.

Peiris, K.H.S., Leffler, R.G., Dull, G.G. & Kays, S.J. (1997). Nondestructive determination of soluble solids content of peach by near infrared

spectroscopy. In: Proceedings from the Sensors for Nondestructive Testing International Conference and Tour, Pp. 77-87. Orlando, Florida.

Rannou, H. & Downey, G. (1997). Discrimination of raw pork, chicken and turkey meat by spectroscopy in visible, near- and mid-infrared ranges. *Analytical Communications*, **34**, 401-404.

Ridgway, C. & Chambers, J. (1996). Detection of external and internal insect infestation in wheat by near-infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture*, **71**, 251-264.

Rodriguez-Otero, J.L., Centena, J. & Hermida, M. (1997a). Application of near infrared transflectance spectroscopy to the analysis of fermented milks. *Milchwissenschaft*, **52**, 196-200.

Rodriguez-Otero, J.L., Hermida, M. & Centena, J. (1997b). Analysis of dairy products by near-infrared spectroscopy: A review. *Journal of Agricultural and Food Chemistry*, **45**, 2815-2819.

Sato, T., Kawano, S. & Iwamoto, M. (1990). Detection of foreign fat adulteration of milk fat by near infrared spectroscopic method. *Journal of Dairy Science*, **73**, 3408-3413.

Sato, T. (1994). Application of principal-component analysis on near-infrared spectroscopic data of vegetable oils for their classification. *Journal of the American Oil Chemists' Society*, **71**, 293-298.

Sato, T., Abe, H., Kawano, S., Ueno, G., Suzuki, K. & Iwamoto, M. (1994). Near-infrared spectroscopic analysis of deterioration indices of soybeans for process control in oil milling plant. *Journal of the American Oil Chemists' Society*, **71**, 1049-1055.

Sato, T., Takahata, Y., Noda, T., Yanagisawa, T., Morishita T. &Sakai, S. (1995). Nondestructive determination of fatty acid composition of husked sunflower (*Helianthus annua L.*) seeds by near-infrared spectroscopy. *Journal of the American Oil Chemists' Society*, **72**, 1177-1183.

Sato, T. (1997). Application of near infrared spectroscopy for the analysis of fatty acid composition. *Lipid Technology*, **9**, 46-49.

Sato, T., Uezono, I., Morishita, T. & Tetsuka, T. (1998). Nondestructive estimation of fatty acid composition in seeds of *Brassica napus L.* by near-infrared spectroscopy. *Journal of the American Oil Chemists' Society*, **75**, 1877-1881.

Scanlon, M.G., Adam, L.R. & Pritchard, M.K. (1997). NIR analysis of dry matter and sugars in whole potato tubers. In: Proceedings from the Sensors for Nondestructive Testing International Conference and Tour, Pp. 88-100. Orlando, Florida.

Schmilovitch, Z., Hoffman, A., Egozi, H., Ben-Zvi, R., Bernstein, Z. & Alchanatis, V. (1997). System and method for pre-harvested maturity determination of fresh dates by near infrared spectrometry. In: Proceedings from the Sensors for Nondestructive Testing International Conference and Tour, Pp. 111-121. Orlando, Florida.

Schmilovitch, Z., Hoffman, A., Egozi, H., Ben-Zvi, R., Bernstein, Z. & Alchanatis, V. (1999). Maturity determination of fresh dates by near intrared spectrometry. *Journal of the Science of Food and Agriculture*, **79**, 86-90.

Schulz, H., Drews, H.H., Quilitzsch, R. & Kruger, H. (1998). Application of near infrared spectroscopy for the quantification of quality parameters in selected vegetables and essential oil plants. *Journal of Near Infrared Spectroscopy*, **6**, A125-A130.

Sirieix, A. & Downey, G. (1997). Commercial wheatflour authentication by discriminant analysis of near infrared reflectance spectra. *Journal of Near Infrared Spectroscopy*, **1**, 187.

Slaughter, D.C. (1995). Nondestructive determinations of internal quality in peaches and nectarines. *American Society of Agricultural Engineers*, **38**, 617-623.

Stark, E., Luchter, K. & Margoshes, M. (1986). Near-infrared analysis (NIRA): A technology for quantitative and qualitative analysis. *Applied Spectroscopy Reviews*, **22**, 342.

Sugden, D. (1997). Post-milling flour processing. *World Grain*, April 1997, 22-26.

Uthayakumaran, S., Wesley, I.J., Bekes, F., Osborne, B.G. & Skerritt, J.H. (1998). Cereals 98 (Edited by L. O'Brien, A.B. Blakeney, A.S. Ross, & C.W. Wrigley) Pp. 79-82. Wringley, RACI, Melbourne, Australia.

Van den Berg, F.W.J., Van Osenbruggen, W.A. & Smilde, A.K. (1997). Process analytical chemistry in the distillation industry using near-infrared spectroscopy. *Process Control and Quality*, **9**, 51-57.

Velasco, L., Pérez-Vich, B. & Fernández-Martìnez, J.M. (1998). *Journal of the American Oil Chemists' Society*, **75**, 1883-1888.

Velasco, L., Goffnam, F.D. & Becker, H.C. (1999). Development of calibration equations to predict oil content and fatty acid composition in *Brassicaceae* germplasm by near-infrared reflectance spectroscopy. *The Journal of the American Oil Chemists' Society*, **76**, 25-30.

Wehling, R.L. (1994). Infrared Spectrometry. In: Introduction to the chemical analysis of foods (edited by S.S. Nielsen). Pp. 344, 346-347, 349. Boston, MA: Jones and Bartlett Publishers.

Wesley, I.J., Gras, P.W., Larsen, N., Osborne, B.G. & Skerritt, J.H. (1998a). Cereals 98 (Edited by L. O'Brien, A.B. Blakeney, A.S. Ross, & C.W. Wrigley) Pp. 443-446. Wringley, RACI, Melbourne, Australia.

Wesley, I.J., Larsen, N., Osborne, B.G. & Skerritt, J.H. (1998b). Non-invasive monitoring of dough mixing by near infrared spectroscopy. *Journal of Cereal Science*, **27**, 61-69.

Westerhaus, M.O. (1989). Calibration: Interpretation of Regression Statistics. In: Near Infrared Reflectance Spectroscopy (NIRS): *Analysis of Forage Quality* (edited by G.C. Marten, J.S. Shenk & F.E. Barton II). Pp. 39-40. Agriculture Handbook No. 643. United States Department of Agriculture: Agriculture Research Service.

Wetzel, D.L. (1983). Near-Infrared reflectance analysis - Sleeper among spectroscopic techniques. *Analytical Chemistry*, **55**, 1165-1171.

Wetzel, D.L.B. (1998). Analytical near infrared spectroscopy. In: Instrumental methods in food and beverage analysis (edited by D.L.B. Wetzel & G. Charalambous). Pp.159, 166-170, 188. Elsevier, New York.

Willard, H.H., Merritt Jr., L.L., Dean, J.A. & Settle Fr., F.A. (1988). Infrared Spectrometry. In: Instrumental methods of analysis, 7th ed. Pp. 302-303. California: Wadsworth Publishing Company.

Williams, P.C. (1991). Prediction of wheat kernel texture in whole grains by near-infrared transmittance. *Cereal Chemistry*, **68**, 112-114.

Williams, P.C. & Stevenson, S.G. (1990). Near- infrared reflectance analyses: Food industry applications. *Trends in Food Science & Technology*, **1**, 44-48.

Zlobsdz, T., Meresz, P. & Salgo, A. (1997). Application of near infrared transmission spectroscopy for the determination of some quality parameters of apples. *Journal of Near Infrared Spectroscopy*, **2**, 213-223.

CHAPTER 3

THE APPLICATION OF FOURIER TRANSFORM NEAR INFRARED (FT-NIR) SPECTROSCOPY IN THE WINE INDUSTRY OF SOUTH AFRICA

Summary

Fourier transform near infrared (FT-NIR) spectroscopy can be used as a rapid method to measure the °Brix content and to discriminate between different must samples in terms of their free amino nitrogen (FAN) values. It can also be used as a rapid method to discriminate between Chardonnay wine samples in terms of the malolactic fermentation (MLF) status. By monitoring the conversion of malic to lactic acid the samples could be classified based on whether MLF has started, is in progress or has been completed. Furthermore, FT-NIR spectroscopy can be used as a rapid method to discriminate between table wine samples in terms of the ethyl carbamate (EC) content. Ethyl carbamate in wine can pose a health threat and needs to be monitored by determining the EC content in relation to the regulatory limits set by authorities. For each of the above-mentioned parameters, QUANT+^{TM} methods were built and calibrations derived and it was found that a very strong correlation existed in the sample set for the FT-NIR spectroscopic predictions for °Brix (r = 0.99, SEP = 0.31%). However, the correlation for the FAN predictions (r = 0.405, SEP = 275%), malic acid (r = 0.64, SEP = 1.02%), lactic acid (r = 0.61, SEP = 1.35%) and EC predictions (r = 0.47, SEP = 3.6%) were not good. The must samples could be classified in terms of their FAN values when soft independent modelling by class analogy (SIMCA) diagnostics and validation were applied as a discriminative method, obtaining results with recognition rates exceeding 80% in all cases. When SIMCA diagnostics and validation were applied to the Chardonnay and EC wine samples, results with recognition rates exceeding 88% and 80% respectively,

were obtained, showing this method as successful to discriminate between samples.

Introduction

Fourier transform near infrared spectroscopy

The NIR spectroscopy method of analysis is an instrumental method for rapid and reproducible measurement of the chemical composition of samples with little or no sample preparation (Norris, 1989). Each of the major chemical components of a food sample has NIR absorption properties, which can be used to differentiate one component from the other. By use of Fourier transform interferometers, FT-NIR diffuse reflectance signals are formed that contain information about the composition of the sample (Willard *et al.*, 1988). Such information can be extracted by appropriate mathematical treatment of the data. NIR spectroscopy is being used for the determination of alcohol content and °Brix in wine and preliminary investigations has been execute on glycosyl-glucose and methanol concentrations in grape spirits where it is usually applied as a correlative technique (Gishen & Dambergs, 1998). This means that a calibration or learning set of samples is analysed by standard laboratory methods as reference and then the same samples are scanned by the spectrophotometer. The data obtained by the reference method are then correlated with the large amount of spectral data, using sophisticated multivariate statistical data analysis techniques, in order to find a correlation that can predict the analytical results from the spectral data. The FT-NIR spectroscopic instrument can then be used to scan new samples to obtain analytical data (Gishen & Dambergs, 1998). A measurement can be made in as little as 10 seconds, although the average would be closer to 30 seconds to 3 minutes. Little or no sample preparation is needed and the technique can be used by employees without extensive training. It is also applicable for on-line measuring systems (Willard *et al.*, 1988, Wehling, 1994).

FT-NIR spectroscopy can also be used for classification and verification of raw materials (Perkin Elmer, 1998a). In many cases where sample classification is applied, it is only necessary to know whether a sample belongs to a specific class or not or whether it is above or below a specific cut-off point. In such cases the data sets are divided into classes to differentiate between the specified properties. FT-NIR was used to discriminate between different tartrates used in the wine industry, classifying the samples as containing calcium tartrate or potassium bitartrate (Perkin Elmer, 1998b). If the materials to be identified are spectroscopically dissimilar, it is often only necessary to use a simple distance measure such as a spectral difference. If the spectra are similar, it may be necessary to include slightly more sophisticated techniques that take into consideration both the variability of the spectra of interest and the differences between the spectra (Perkin Elmer, 1998a). The soft independent modelling by class analogy (SIMCA) technique provides such a method.

When building the SIMCA method, a principal component model is created for each class. An "envelope" to contain the standards of a class is constructed. If one principal component is used, the class mean and a line can model the data. If two principal components are used, the class mean and a plane are required, and with three principal components, the class mean and a volume are required (Perkin Elmer, 1997). If there are more than three components, the envelope can be thought of as a box with more than three dimensions (a hyperbox). In a SIMCA classification the unknown spectrum is classified by whether it lies inside or outside the hyperbox (this is the model residual). SIMCA is thus a method that provides a set of parameters that characterise each class and are the basis for other quantities that describe the data.

Fermentation and the optimal nitrogen balance of must

The nitrogen content of grapes impacts the production of yeast biomass, fermentation rate and time to completion of a fermentation, and can influence the spectrum of end products of yeast metabolism (Bisson, 1991). There are three major nitrogen-containing classes of compounds in yeast: the amino acids and

their derivatives; the nucleotide bases and their derivatives; and the polyamines and grape juice contains varying concentrations of all these components (Bisson, 1991).

The total nitrogen content and distribution among these compounds are highly variable in grapes (Amerine *et al.*, 1980). The total level of free amino acids in musts ranges from 65 to 1130 $mg.L^{-1}$. This variation is dependent upon grape variety, growing region, time of harvest, as well as being correlated with the nitrogen richness of the soil and with fertilization practices. A value of 500 $mg.L^{-1}$ of nitrogen in must was reported as necessary to achieve maximal yeast biomass production (Agenbach, 1978). In addition to an impact of nitrogen content on cell production, nitrogen also affects the fermentation rate. At least 140 $mg.L^{-1}$ of assimilable nitrogen is needed in juice or must in order for the yeast to complete fermentation to dryness (Agenbach, 1978).

Free amino (or alpha) nitrogen (FAN) has often been utilized as an indicator of the nitrogen richness or nitrogen availability for yeast growth and fermentation (Amerine & Ough, 1980). Statistical analyses established the FAN/°Brix ratio as the most reliable means of determining optimal nitrogen balances in must (Vos *et al.*, 1979).

The natural FAN content of musts from mature grapes of most cultivars (Pinotage is the exception), ranges from approximately 400 to 1000 $mg.L^{-1}$ N when ammonium sulphate is used as reference standard. With must samples at lower levels, the addition of a maximum of 500 $mg.L^{-1}$ N would thus ensure a total FAN content of at least 800 $mg.L^{-1}$ N, the minimum concentration required for maximum fermentation rates (Vos *et al.*, 1980). Highly significant exponential correlations exist between the FAN content of must and fermentation rate as well as yeast mass production (Agenbach, 1978). In other words, when FAN is no longer a limiting factor, subsequent increases of assimilable nitrogen will not cause further stimulation of total yeast mass production and fermentation rate.

Variations in the FAN content would reflect changes in concentrations of free amino acids, low-molecular-weight peptide concentrations and ammonia. Therefore, the FAN content should be an accurate index of the nitrogen

requirements of yeast and hence fermentation rates. The FAN/°Brix ratio is now established as a superior index and indicates that FAN requirements of yeast are influenced mainly by the sugar content of the musts, but also by the temperature of fermentation and the species of yeast used. This is also in agreement with the results of Agenbach where increases in sugar concentration resulted in larger consumption of assimilable nitrogen (Vos *et al.*, 1980).

The phenomenon of lagging or stuck fermentations represents one of the most annoying and troublesome problems experienced during wine making (Vos *et al.*, 1980). Stuck fermentations are those where the final phases of vinification are protracted and lingering and where the completion of the fermentation is far from certain (Kunkee, 1991). It is at the end phase of such vinifications when the diminished metabolic activity of the yeast brings about a diminished production of the protective carbon dioxide gaseous blanket, and when the wine is especially susceptible to the peril of oxidation. Furthermore, the finishing operations required for cellar storage cannot be commenced until the fermentation is completed. Thus, sluggish fermentations are precarious with respect to the quality of the end product and costly with respect to demand for attention. Additionally, the longer fermentation times result in less turnover of fermentation cooperage, often an economic concern (Kunkee, 1991).

The causes of sluggish fermentations seem to be several, however, there is a direct relationship between end phase fermentation rates and the nitrogen content of the must. The most apparent, and certainly the most studied influence on the fermentation is, of course, the nitrogenous components. Some relationships between the amount of assimilable nitrogen, either naturally present, reflecting climate, vineyard, and fertilization treatments, or by nutrient supplementation to the must have been reviewed (Kunkee, 1991).

Malolactic fermentation in wine

Most red wines and some white wines in colder wine regions are subjected to the secondary malolactic fermentation (MLF) during or soon after alcoholic fermentation (Volschenk *et al.*, 1997). Malolactic fermentation in wine is caused

by the metabolic activity of certain lactic acid bacteria (LAB), where the most important aspect is the microbial deacidification which results from the decarboxylation of malic acid to lactic acid (Nielsen *et al.*, 1996). Total acidity decreases and pH increases, resulting in wine with a softer palate (mouthfeel). Malolactic fermentation also contributes to the flavour and complexity, and it increases the microbiological stability of the wine (Nielsen *et al.*, 1996).

Malolactic fermentation may occur spontaneously in wine as a result of the growth of an indigenous flora of lactic acid bacteria originating from the vines and grape skins and also surviving on winery equipment. When occurring in this fashion, MLF is often delayed and may take place several months after the alcoholic fermentation (Nielsen *et al.*, 1996).

The composition of the wine directly affects the ability of strains to carry out MLF (Vaillant *et al.*, 1995). The malolactic activity of bacteria and their growth depend on various factors: in particular pH; ethanol; SO_2 content; and temperature. Other factors such as amino acids, sugars and organic acids also interfere.

Malolactic fermentation plays three important roles for the winemaker: first, deacidification of wine through release of carbon dioxide, which is due to loss of a carboxyl group during the conversion of malic acid to lactic acid (Figure 1); second, bacterial stabilisation by utilization of the micronutrients in wine during growth of these nutritionally fastidious microbes; and third, formation of flavourful end products. Each of these roles affects the final odiferous properties of wine (Kunkee, 1997).

Malolactic fermentation can be considered a spoilage factor if it takes place under the wrong conditions such as grapes from warm regions which have too little malic acid to begin with; wines to which microbial stabilization agents have not been carefully added; and wines where the end products can be too flavourful, simply unpleasant, or worse (Kunkee, 1997).

A major disadvantage of MLF is the unpredictability of its occurrence and control during vinification (Volschenk *et al.*, 1997). Spontaneous MLF may occur

during alcoholic fermentation or only months after, and even the application of bacterial starter cultures does not completely ensure rapid MLF.

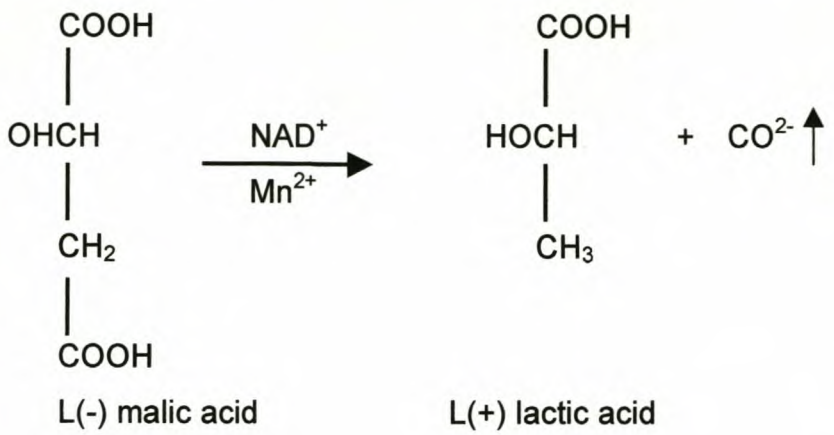


Figure 1. Malolactic fermentation transferring malic acid to lactic acid and allowing deacidification of wine (Kunkee, 1997).

The implication of ethyl carbamate for the wine industry

Ethyl carbamate (EC) or urethane is a naturally occurring component in all fermented foods and beverages. Because EC has shown a potential for carcinogenity when administered in high doses in animal tests, the wine industry have to monitor the EC levels in their products (Butzke & Bisson, 1997). Ethyl carbamate is not an intentionally added substance, it forms during fermentation of alcoholic beverages, and, if the fermented product is heated, such as in "baking" sherry or distilling spirits, its levels increase (Segal, 1988).

Urea, a natural by-product of yeast metabolism, is the main precursor of ethyl carbamate in wines (Figure 2) (Monteiro *et al.*, 1989). Arginine, along with proline, are generally the major amino acids found in grape juice. The enzyme arginase catalyses the cleavage of arginine to ornithine and urea. All known yeasts possess this enzymatic activity. The resulting urea can also be used as a nitrogen source and further broken down to ammonia and carbon dioxide by the yeast *Saccharomyces cerevisiae.* This takes place via a degradative enzyme complex, composed of urea carboxylase and allophanate hydrolase (Henschke &

Ough, 1993). However, this process may not be complete before the end of fermentation if the must originally contained high levels of nitrogenous compounds (*i.e.* high alpha-amino acids) that are metabolized by yeast before arginine and urea. Residual levels of urea remaining after fermentation can react with EC. This reaction is temperature and time dependent (Henschke & Ough, 1993).

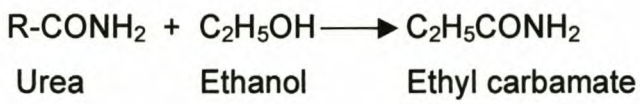


Figure 2. The conversion of urea and ethanol to form ethyl carbamate (Monteiro *et al.*, 1989).

Concern in the USA over EC began in November 1985 with news reports that Canadian authorities had detected the chemical in certain wines and distilled spirits (Segal, 1988). At that time, the US Food and Drug Administration (FDA) and the Bureau of Alcohol, Tobacco, and Firearms (BATF) began carefully evaluated studies to determine whether there could be a long-term health risk to consumers from EC in alcoholic beverages. In December 1985, Canada set regulatory limits for table wine at 30 $\mu g.kg^{-1}$; ports and sherries at 100 $\mu g.kg^{-1}$; distilled spirits at 150 $\mu g.kg^{-1}$; and fruit brandies, cordials and liqueurs at 400 $\mu g.kg^{-1}$ (Segal, 1988). In December 1987 the FDA accepted a plan of the Distilled Spirits Council of the United States to reduce the levels of EC in whiskey to 125 $\mu g.kg^{-1}$ or less. In January 1988, FDA also accepted an EC reduction plan for table and dessert wines from the major U.S. wine producers, represented by the Wine Institute and the Association of American Vintners. The plan called for EC levels to average no more than 15 $\mu g.kg^{-1}$ in table wines. For dessert wines, such as cream sherries, which contain more than 14% alcohol, EC levels were to average no more than 60 $\mu g.kg^{-1}$ (Segal, 1988). Thus, the FDA proposed to reduce EC content to the above voluntary targets.

In South Africa no regulatory limits for EC levels in wines exists, but wines that are exported to countries with regulatory limits have to show the EC content (M. Waldner, ARC Infruitec-Nietvoorbij, personal communication). It is therefore necessary to monitor the EC content in some wines to determine whether they fall inside the regulatory boundaries of these countries.

Factors influencing the formation of EC are: temperature and time of storing, which is considered to be the most important factor affecting the rate of EC formation for EC concentrations increase with increased temperatures (Stevens & Ough, 1993); malolactic fermentation (Liu *et al.*, 1994); vineyard fertilization (Ough *et al.*, 1989); and aeration (Henschke & Ough, 1991). A complete elimination of EC is, however not possible (Butzke & Bisson, 1997).

Objective

Currently the FAN, malic and lactic acid and EC measurements are monitored using expensive, quantitative, time-consuming analytical methods, such as GC and HPLC, whereas FT-NIR spectroscopy can be used as a rapid method that requires no sample preparation. Although the measurement of the sugar content by use of a Balling meter is a simple and fast method, simultaneous determination of the FAN and °Brix values will be time-saving.

The objective of this study was to develop FT-NIR spectroscopy methods for the determination of FAN and °Brix values of must samples, malic and lactic acid values of Chardonnay samples undergoing MLF and EC values of white and red table wines. A classification technique, SIMCA, has also been used on FT-NIR spectra of must and wine samples to discriminate between the samples in terms of their FAN values, malic and lactic acid and EC content.

Materials and methods

A. Sugar content and FAN value

Wine samples

A selection of 97 must samples of white grape varieties, representative of the Western Cape region, was drawn from settling tanks at the cellars of Distillers Corporation in Stellenbosch, South Africa. The set included the following samples: 46 Steen; 29 Sauvignon Blanc; 9 Chardonnay; 9 Riesling; 5 Pinot Noir; and 2 Gewurtztraminer. The must samples were collected after one day in the settling tanks followed by a few hours of cold storage during the harvest period over 2 consecutive seasons (1999 & 2000) and analysed on receipt without any pre-conditioning.

Chemical analyses

The FAN content of the must samples was determined by means of a spectrophotometer equipped with an auto-analyser (Vos, 1977). The sugar content ($^{\circ}$Brix) of the must samples was determined by means of a Balling meter.

Fourier transform near infrared (FT-NIR) spectroscopy measurements

Fourier transform near infrared (FT-NIR) spectroscopy analyses of the grape must were carried out in transmission mode. The spectra were recorded in a 0.5 mm quartz cuvette at 4 cm^{-1} intervals with a 8 scan sequence, using a Spectrum IdentiCheck^{TM} 2.0 FT-NIR System (Perkin Elmer). The wavelength region for all calibrations was 10 000 to 4000 cm^{-1} (1000 to 2500 nm) resulting in a total of 1501 data points per spectrum.

Data analyses:

Data manipulation: PLS regression

Multiplicative scatter correction (MSC) was applied to the spectra to eliminate interference of scatter, and then transformed with second derivative processing. Pretreatment and calibration model development was performed using QUANT+^{TM} 4.1 software (Perkin Elmer). The partial least squares (PLS) algorithm was used to derive calibrations to predict FAN content and ^{o}Brix in wine samples. Partial least squares can be described as a projection of the NIR

spectral data and the chemical data onto a few latent orthogonal factors, retaining the main part of the information for both spectral and chemical data (Garcia-Jares & Medina, 1997). This results in reduced spectral data without discarding useful information (Osborne *et al.*, 1993). In the PLS model, both the independent (spectral data) and dependent (chemical data) variables participate in the construction of the latent variables. The latent variables of the independent set of data not only represent the original data, but are also correlated to the dependent set of data by its latent variables. Partial least square achieves a compromise between the explanation of the spectral variables and the prediction of the chemical variables.

Cross validation was performed to construct PLS factors from the original spectral data. Cross validation (Osborne *et al.*, 1993) is a technique where the calibration samples are divided into a number of groups, and the prediction of samples within a group is based on a calibration made using samples from all other groups. In this study cross validation was performed by taking one sample from the calibration set and that sample is being predicted by the calibration model using the remaining samples. This procedure was repeated until all samples were predicted once.

The accuracy of the calibrations was expressed as the standard error of cross validation (SECV) of the bias-corrected residuals (p. 26). The bias (p. 25) is interpreted as the average difference between y and $\hat{y}_i$ in the prediction set. If the BIAS is near a value of zero, the overall error of validation can be interpreted as the standard deviation (SD) of the NIR prediction.

Alternatively the accuracy was calculated as the root mean square standard error of cross validation (RMSECV) (p.25), which is preferred in many cases (Næs & Isaksson, 1991). This is because the removal of the bias-effect (as done in the SECV determination) can give an over optimistic impression of the prediction error when the equation is used for future samples which are different from those used for validation. The reason for this is that the bias for a limited set of samples has a strong random component that can change from one validation set to another.

Upon completion of the calibration, the model was validated with an independent set of samples. The spectra were randomly divided into two sets: 79 samples (ca. 70%) were used for the calibration set and 30 samples (ca. 30%) for the validation set. The accuracy of the calibration was expressed as the standard error of prediction (SEP) of the bias-corrected residuals (p. 25). Alternatively the accuracy of the calibration when predicting an independent set of samples was expressed as the root mean square error of prediction (RMSEP) (p. 24). RMSEP is an estimate of the accuracy of the calibration against the reference method, and is calculated using an independent test set.

The standard deviation of the reference data (SD) divided by the SEP or SECV is called the standard deviation of reference data (RPD) (equation 6) (Williams, 1991). The RPD is an indication of the efficiency of a calibration (pp. 26-27).

SIMCA classification

The spectra were divided into Class 1 (1 – 800 $mg.L^{-1}$ N) where it might be necessary to add extra nitrogen for a complete fermentation and Class 2 (800 – 2000 $mg.L^{-1}$ N) where enough nitrogen is present to complete the fermentation. Principal component analyses (PCA) models were derived for the two classes and SIMCA models were created to allow differentiation between the classes. The validation set was comprised by selecting samples from the two classes prior to SIMCA model building and consisted of 12 samples. After diagnostics was performed on the SIMCA models, the validation set was predicted by each of the models and decisions on their affiliation were made based on their distance from the nearest cluster model.

B. Malolactic fermentation

Wine samples

A selection of 65 Chardonnay wine samples were drawn from barrels at the cellars of Distillers Corporation in Stellenbosch, South Africa and another 43

Chardonnay samples were received from the department of Wine Biotechnology at the University of Stellenbosch, South Africa. The samples, stored at 4°C, were collected over a 3 month period and each sample analysed once on receipt.

Chemical analyses

The malic and lactic acid content of the wine samples was determined by means of high-pressure liquid chromatography (HPLC) (Schneider *et al.*, 1987).

Fourier transform near infrared (FT-NIR) spectroscopy measurements

Fourier transform near infrared (FT-NIR) spectroscopy analyses of the wine samples were carried out in transmission mode as described for sugar content and FAN value, except that a 16 scan sequence was used.

SIMCA classification

Principal component analyses (PCA) models were developed and three SIMCA models were created using PCA models of the classes. Class 1 (0 - 0.3 $g.L^{-1}$) represents the samples where MLF have not started, Class 2 (0.3 - 2 $g.L^{-1}$) where MLF is underway and Class 3 (> 3 $g.L^{-1}$) where MLF have been completed. The accuracy of the SIMCA models were determined using an independent validation set to perform future classification of unknown samples. The validation set consisted of 22 samples and was predicted by each of the models.

C. Ethyl carbamate

Wine samples

A selection of 200 wine samples was drawn from barrels at the cellars of the ARC Infruitec-Nietvoorbij in Stellenbosch, South Africa. The samples were collected over a period of 2 months (February 1999 and January 2000) and analysed on receipt.

Chemical analyses

The EC content of the wine samples was determined by means of gas chromatography with mass selective detection (GC/MS) according to the OIV method (Canas *et al.*, 1994).

Fourier transform near infrared (FT-NIR) spectroscopy measurements

Fourier transform near infrared (FT-NIR) spectroscopy analyses of the wine samples was carried out in transmission mode as described for sugar content and FAN value, except that a 16 scan sequence was once again used.

SIMCA classification

Partial least squares regression was used to calibrate the FT-NIR spectral data against the laboratory data. The spectra were then classified into Class 1 (0 - 10 $\mu g.kg^{-1}$), Class 2 (10 - 15 $\mu g.kg^{-1}$) and Class 3 (>15 $\mu g.kg^{-1}$) based on the EC values of the samples. Class 1 represents the samples where the EC content possess no threat as health risk, Class 2 where the EC content is close to the restricted value and should be tested to determine the exact EC value and Class 3 that contains the samples where the EC content is above the restricted values. Three SIMCA models were created from PCA models of the different classes and validated on the validation set consisted of 10 samples after diagnostics have been performed on the SIMCA models.

Results and discussion

A. Sugar content and FAN

It was found that a very strong correlation existed in the sample set (combined seasons: 1999 & 2000) for the FT-NIR spectroscopic predictions of °Brix in the must (r = 0.99, SECV = 0.31%; r = 0.99, SEP = 0.31%) (Table 1, Figure 3 & 4). The strong correlation for °Brix was expected, given that the measurement of

°Brix in grape juices by NIR spectroscopy has been well-established (Gishen & Dambergs, 1998). Very good and good RPD values were obtained for the two validation sets with respective values of 6.68 and 5.95.

Calibrations established for the FT-NIR spectroscopic prediction of the FAN content in the must were not as accurate (r = 0.62, SECV = 272.1%; r = 0.405, SEP = 275%) (Table 1, Figure 5 & 6). The respective RPD values were 0.83 and 1.1 for the respective validation sets which is not recommended in terms of classification.

The calibration conditions of the must samples representing only the 1999 season are summarised in Table 2. By comparing the conditions of the combined °Brix calibration set (1999 and 2000) and that of the 1999 °Brix calibration set, it can be seen (RPD = 7.5, SEP = 0.27% for 1999 and RPD = 5.95, SEP = 0.31% for the combined set) that great differences do not exists. The same tendency is also shown when the combined FAN calibration set and that of the 1999 FAN calibration set is compared (RPD = 0.79, SEP = 236.8% for 1999 and RPD = 1.1, SEP = 275% for the combined set). This indicates that different seasons do not influence a calibration to a great extent.

Residual variance plots of the SEP values versus the number of principal components (PC's) used in the °Brix and FAN calibration models of the combined season (Figure 7), illustrate the reason behind the choice of the number of factors. The number of factors used in a calibration must be chosen with caution to prevent overfitting, and therefore the sample set's size and the trend line of the SEP plot must be kept in consideration. In the case of the °Brix calibration, 5 PC's were chosen and 2 in the case of the FAN calibration (Figure 7).

As a result of the poor calibration obtained with the FAN values, SIMCA classification were applied on the FAN data. The two models that were created (class 1 with FAN values between 1 and 800 $mg.L^{-1}$ N and class 2 with FAN values between 800 and 2000 $mg.L^{-1}$ N) showed good classification possibilities. The procedure checks every standard spectrum to ensure that the spectra from a single class fit that class (recognition) and that those from other classes selected are rejected (rejection). For both the data sets the recognition rates were above

87% (Class 1 = 100%, Class 2 = 87%) indicating good separation of each class (Figure 7).

Table 2. Summary of the calibration results of the combined 1999 & 2000 seasons' data sets.

	Full cross validation			Independent validation	
	°Brix	FAN ($g.L^{-1}$)		°Brix	FAN ($g.L^{-1}$)
Range	17 - 27	590 - 2100	**Range**	17 - 27	590 - 2100
Mean	21.62	1271	**Mean**	21.54	1217
SECV	0.306	272.1	**SEP**	0.31	275
RMSECV	0.318	346	**RMSEP**	0.343	351
BIAS	0.004	-1.559	**BIAS**	0.128	52.73
r	0.99	0.602	**r**	0.99	0.405
n	129	123	**n (calibr.)**	84	100
			n (indep.)	43	52
PLS factors	6	5	**PLS factors**	5	2
SD	2.04	334.5	**SD**	2.04	324.3
RPD	6.68	0.83	**RPD**	5.95	1.1

Table 3. Summary of the validated calibration's statistical results of the must samples of the 1999 season.

	Full cross validation			Independent validation	
	°Brix	FAN ($g.L^{-1}$)		°Brix	FAN ($g.L^{-1}$)
Range	17 - 27	590 - 2100	**Range**	17 - 27	590 - 1890
Mean	21.25	1162	**Mean**	21.25	1162
SECV	0.26	252	**SEP**	0.27	236.8
RMSECV	0.27	257	**RMSEP**	0.307	245
BIAS	-0.05	3.97	**BIAS**	0.05	-30.6
r	0.99	0.55	**r**	0.99	0.56
n	97	97	**n (calibr.)**	65	65
			n (indep.)	32	32
PLS factors	3	2	**PLS factors**	4	1
SD	2.024	298.05	**SD**	2.024	298.05
RPD	7.78	1.16	**RPD**	7.5	0.079

Following the results from the diagnostic procedure, the two models were validated. This procedure validates the methods (data divided into classes) that have been built using test spectra that were removed from the data sets before the PCA models were built, i.e. a validation of the independent spectra. Again good results were obtained with recognition rates above 88% (class 1 = 88%, class 2 = 100%) indicating successful classification (Figure 8).

B. Malolactic fermentation

The calibration obtained for the prediction of malic (r = 0.58, SECV = 1.06%) (Table 3 & Figure 9) and lactic (r = 0.51, SECV = 1.14%) (Table 3 & Figure 10) acid did not perform to an acceptable accuracy. For the independent validation sample set, similar results were obtained for the malic (r = 0.64, SEP = 1.024%) (Table 3 & Figure 11) and for the lactic acid (r = 0.61, SEP = 1.35%) (Table 3 & Figure 12). The respective RPD values were 1.19 and 1.13 for the malic acid calibrations and 0.96 and 1.13 for the lactic acid calibrations, confirming the inaccuracy of the quantitative calibrations.

Currently the status of the MLF is determined by means of quantitative analysis. As it is only necessary to know whether the MLF has started, is in progress or has been completed, SIMCA methods have been built. With three models created (Class 1 with lactic acid values between 0 and 0.3 $g.L^{-1}$, Class 2 with values between 0.3 and 2 $g.L^{-1}$ and Class 3 with values above 3) accurate classifications were possible.

Recognition rates of above 95% were reported, indicating good classification of each class (Figure 13). Following the diagnostic procedure, the three data sets were validated using a validation set with independent spectra. Again good results were obtained with the recognition rates reported above 88% (Classes 1 & 2 = 100%, Class 3 = 88%), indicating good separation (Figure 13).

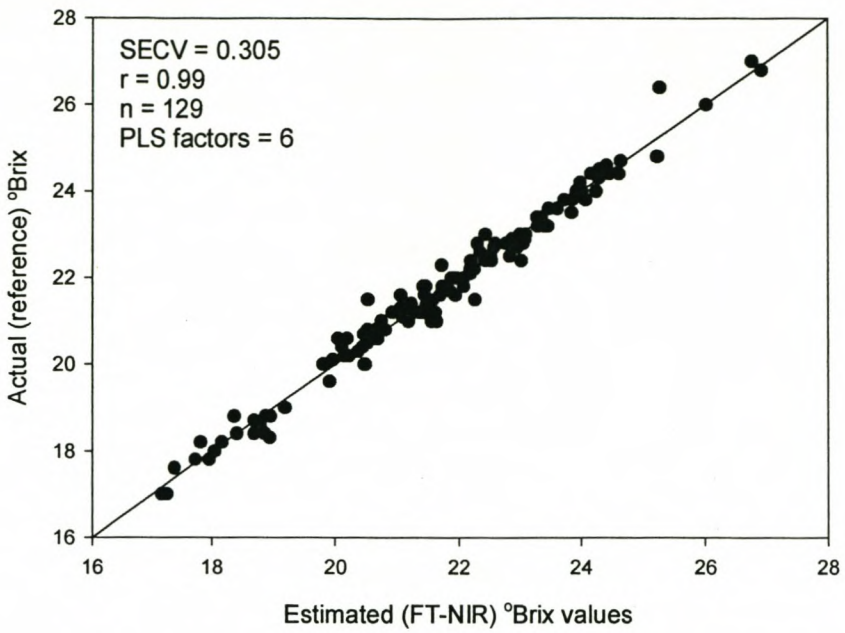


Figure 3. A plot of the estimated (FT-NIR) °Brix values versus actual (physical analyses) °Brix values for the calibration model on must samples.

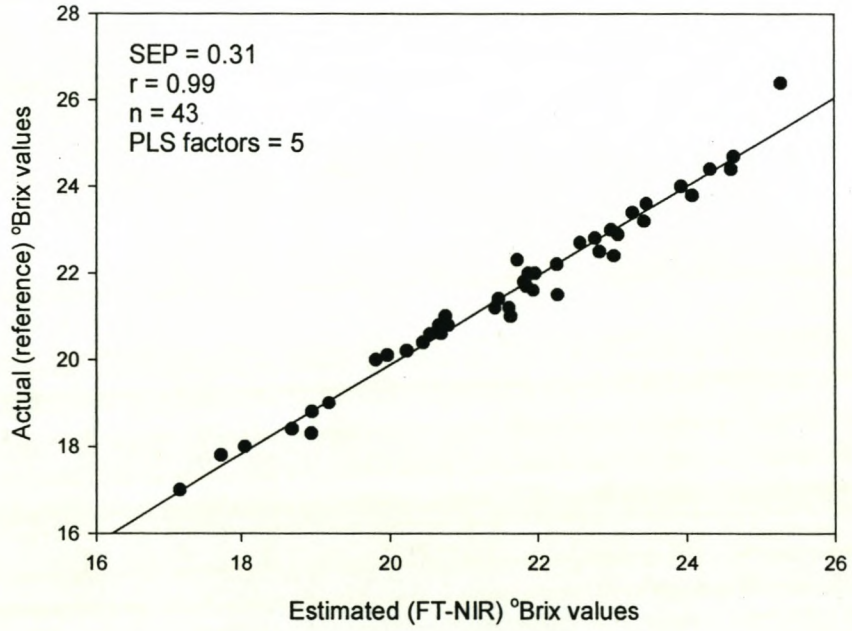


Figure 4. A plot of the estimated (FT-NIR) °Brix values versus actual (physical analyses) °Brix values of the validation must samples.

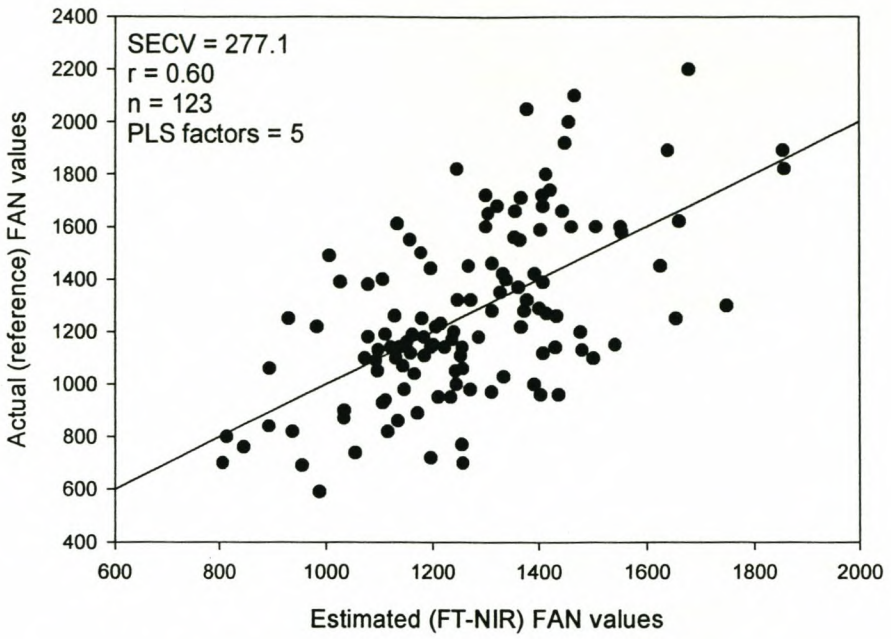


Figure 5. A plot of the estimated (FT-NIR) FAN values versus actual (chemical analyses) FAN values for the calibration model on must samples.

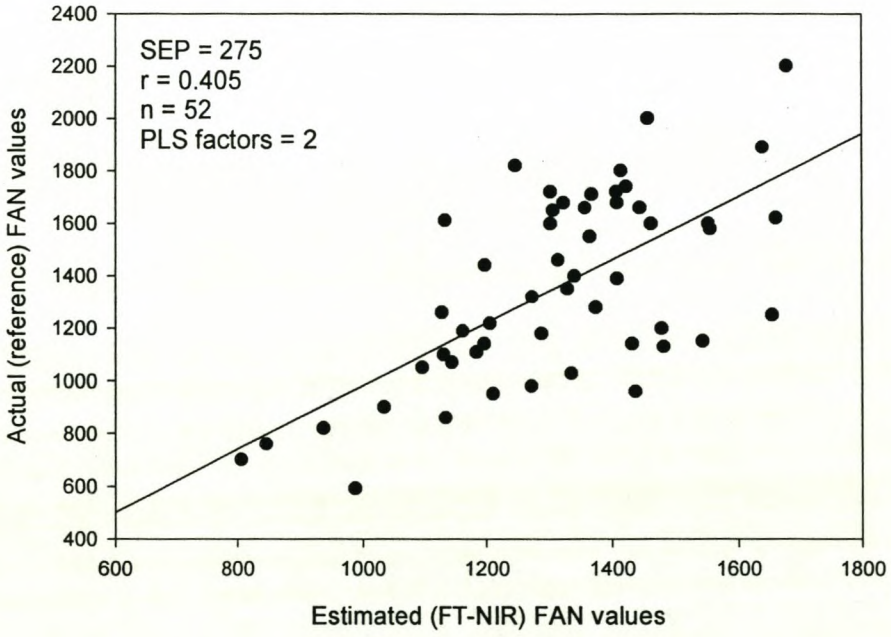
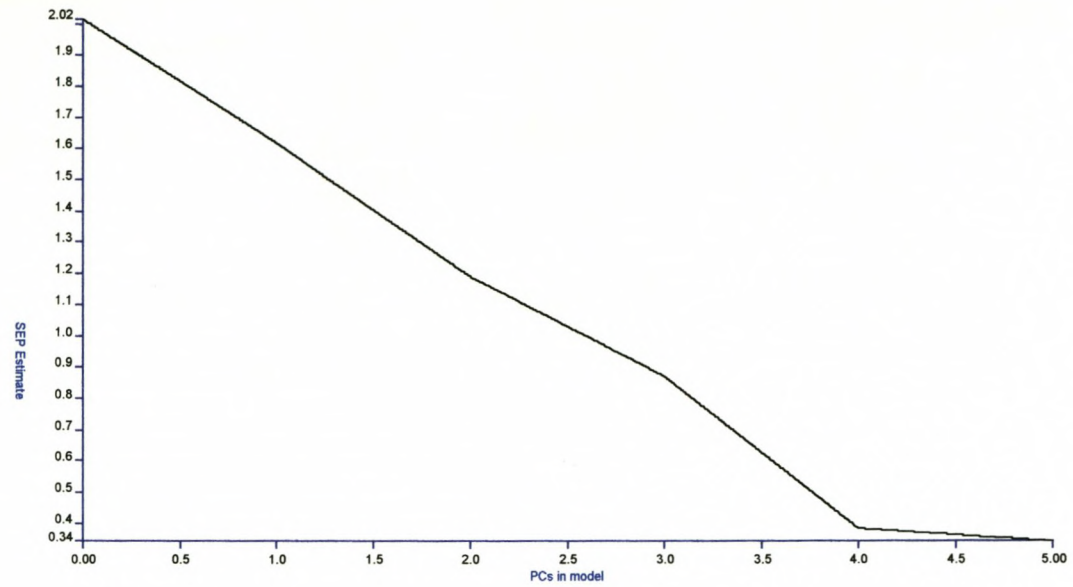
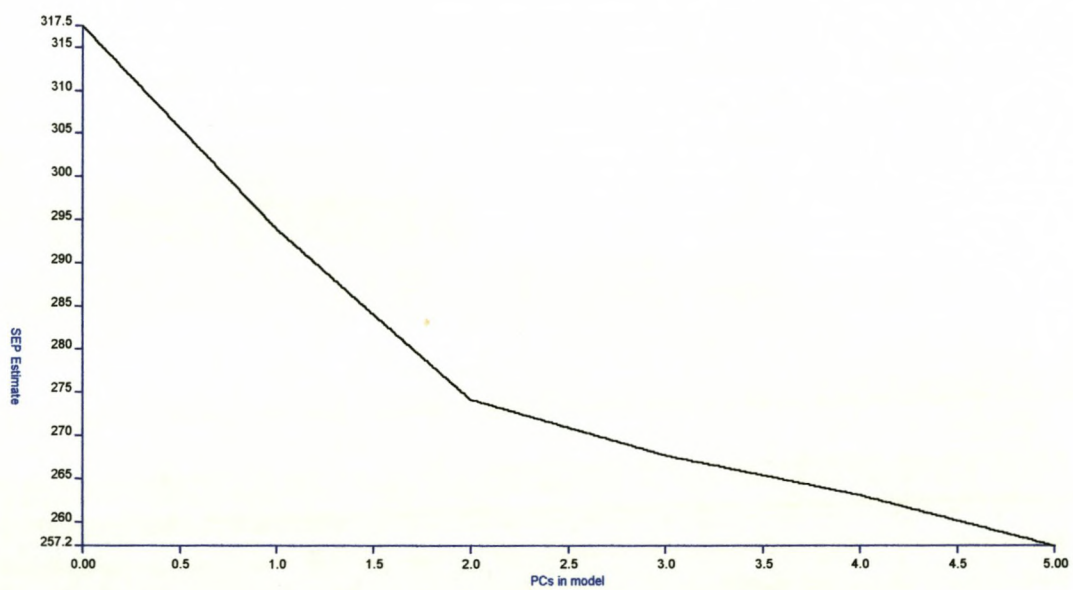


Figure 6. A plot of the estimated (FT-NIR) FAN values versus actual (chemical analyses) FAN values of the validation must samples.



(a)



(b)

Figure 7. The residual variance plots of the SEP values versus the number of principal components used for the combined season's ºBrix (a) and FAN (b) calibration models.

Figure 8. Graphic representation of SIMCA diagnostic and validation results for FAN classification.

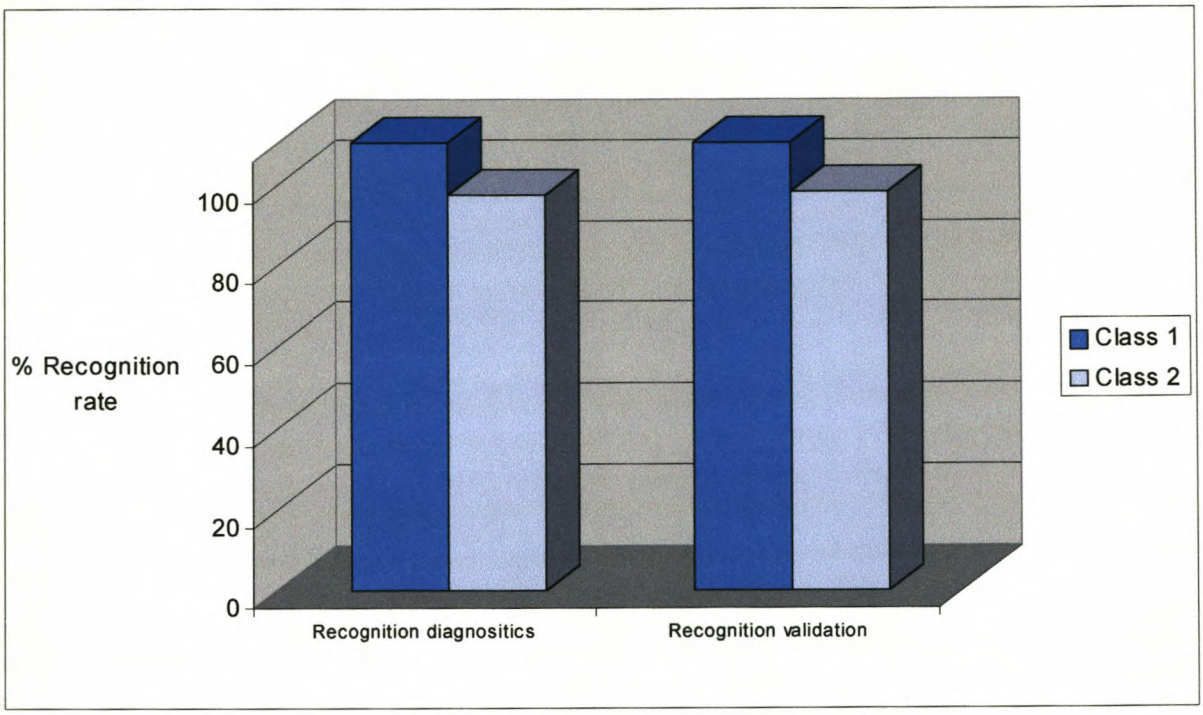


Table 3. Summary of the validated calibrations' statistical results of the wine samples undergoing MLF.

Full cross validation			Independent validation		
	Malic acid ($g.L^{-1}$)	**Lactic acid** ($g.L^{-1}$)		**Malic acid** ($g.L^{-1}$)	**Lactic acid** ($g.L^{-1}$)
Range	0 - 4.78	0 - 5.62	**Range**	0 - 4.78	0 - 5.62
Mean	1.158	1.856	**Mean**	1.158	1.856
SECV	1.055	1.139	**SEP**	1.024	1.345
RMSECV	0.992	1.104	**RMSEP**	0.967	1.102
BIAS	0.063	0.035	**BIAS**	0.027	0.243
r	0.585	0.513	**r**	0.636	0.608
n	109	103	**n (calibr.)**	73	73
			n (indep.)	36	36
PLS factors	4	3	**PLS factors**	4	4
SD	1.2	1.282	**SD**	1.217	1.292
RPD	1.138	1.126	**RPD**	1.188	0.961

C. Ethyl carbamate

Correlation, not accurate enough for quantitative predictions, existed in the sample sets for the FT- NIR spectroscopic predictions of EC (r = 0.55, SECV = 3.67%; r = 0.47, SEP = 3.60%) (Table 4, Figures 14 &15). RPD values of 0.96 for the full cross validation set and 1.06 for the independent validation set, confirmed the inaccuracy of the quantitative calibrations (p. 26).

As a result of the poor calibration obtained with the EC data sets, SIMCA classification diagnostics were applied. The three models that were created (Class 1 with EC values between 0 and 9.99 $\mu g.kg^{-1}$, Class 2 between 10 and 15 $\mu g.kg^{-1}$ and Class 3 with values above 15 $\mu g.kg^{-1}$) showed good classification possibilities.

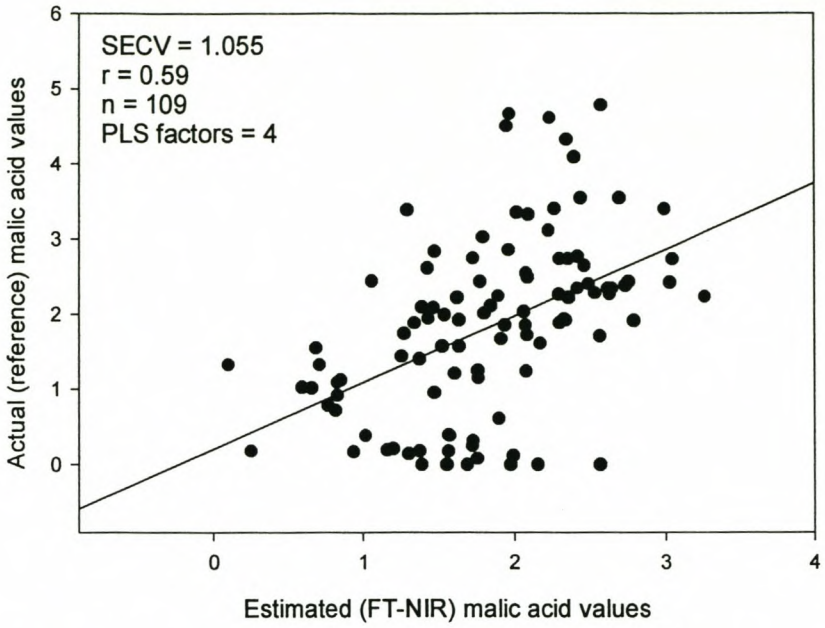


Figure 9. A plot of the estimated (FT-NIR) malic acid values versus actual (chemical analyses) malic acid values for the calibration model on wine samples.

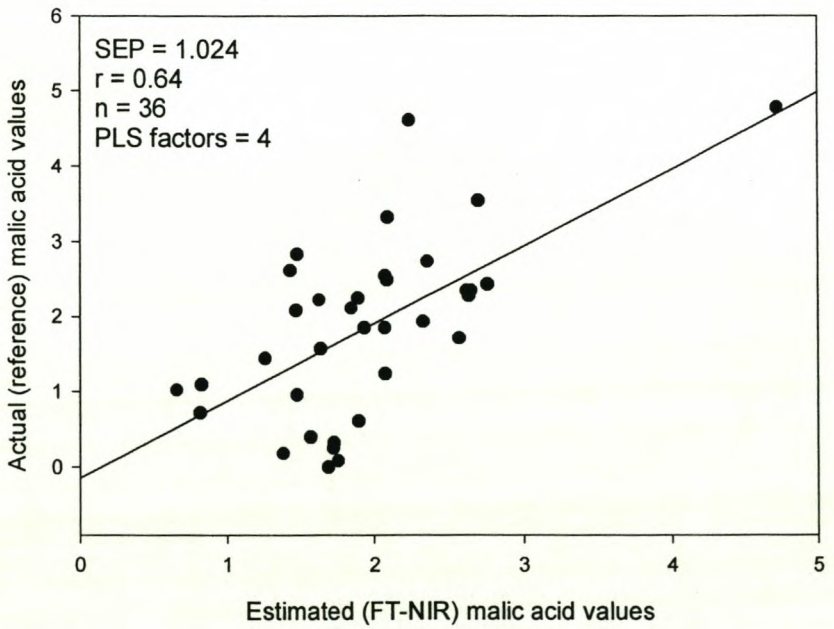


Figure 10. A plot of the estimated (FT-NIR) malic acid values versus actual (chemical analyses) malic acid values of the validation wine samples.

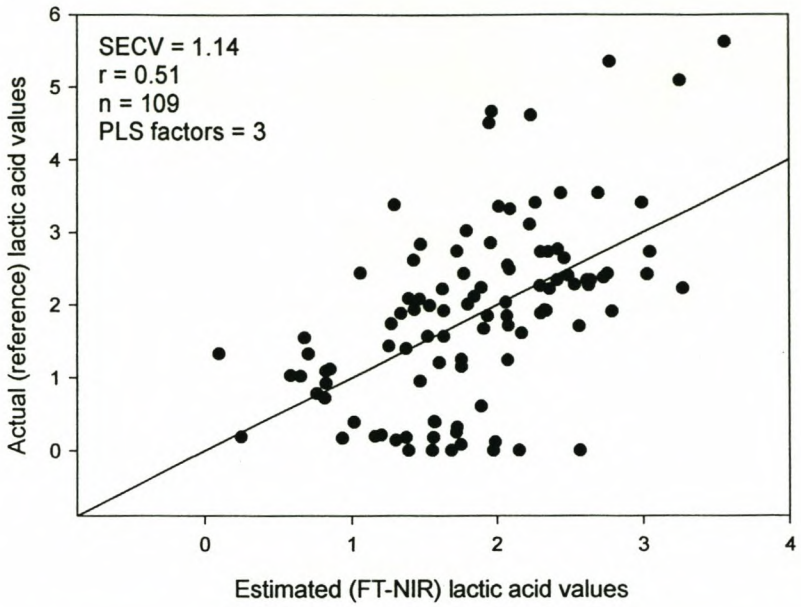


Figure 11. A plot of the estimated (FT-NIR) lactic acid values versus actual (chemical analyses) lactic acid values for the calibration model on wine samples.

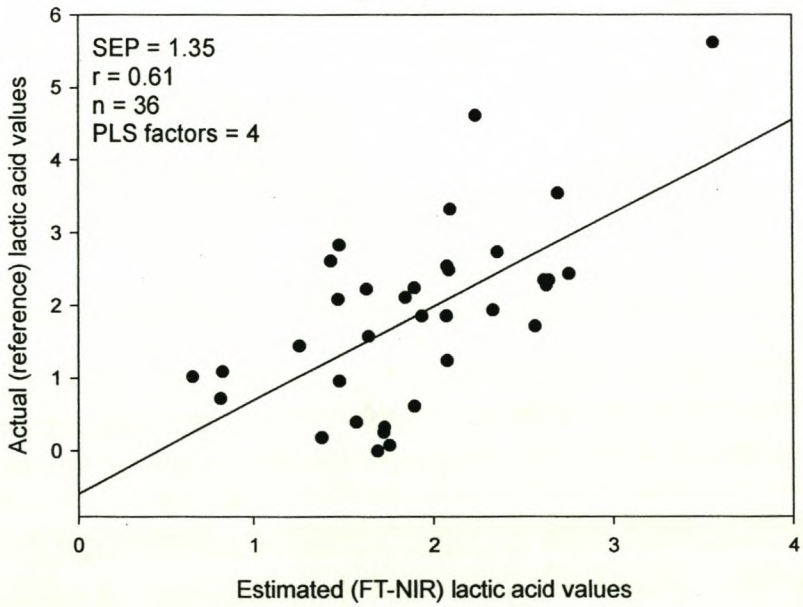


Figure 12. A plot of the estimated (FT-NIR) lactic acid values versus actual (chemical analyses) lactic acid values of the validation wine samples.

Figure 13. Graphic representation of SIMCA diagnostic and validation results.

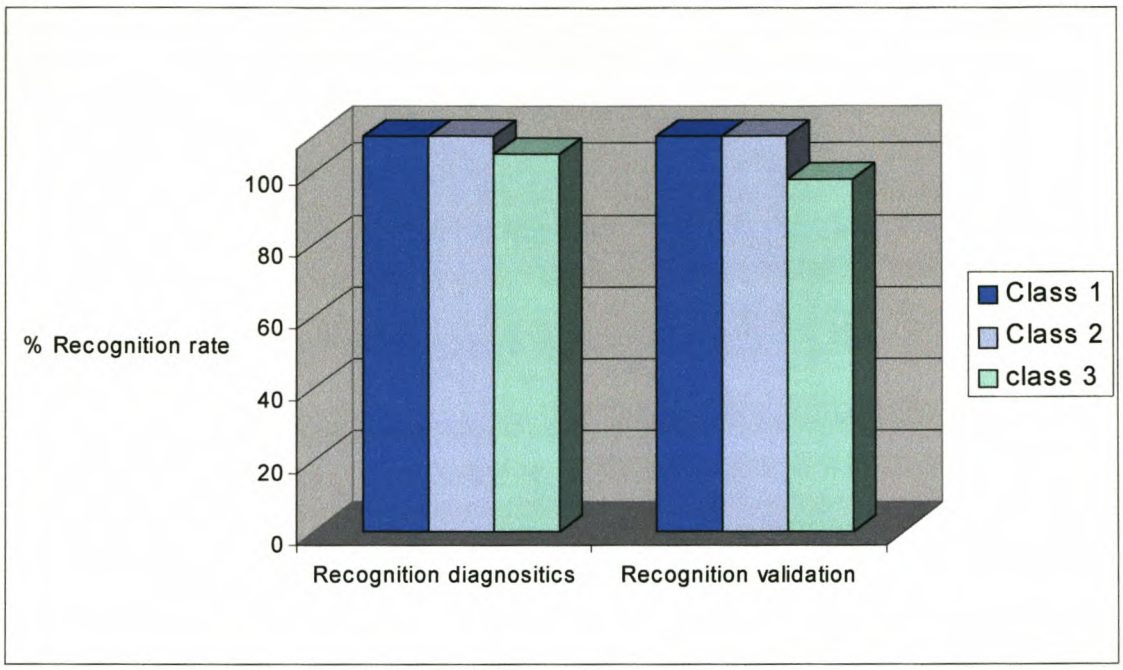


Table 4. Summary of the validated calibrations' statistical results of the table wine samples.

	Full cross validation		**Independent validation**
	EC (µg.kg^{-1})		**EC** (µg.kg^{-1})
Range	0.41 - 19.30	**Range**	0.41 - 19.30
Mean	6.133	**Mean**	5.85
SECV	3.67	**SEP**	3.6
RMSECV	3.62	**RMSEP**	3.51
r	0.55	**r**	0.47
n	156	**n (calibr.)**	56
		n (indep.)	115
PLS factors	3	**PLS factors**	4
SD	3.83	**SD**	3.79
RPD	0.96	**RPD**	1.06

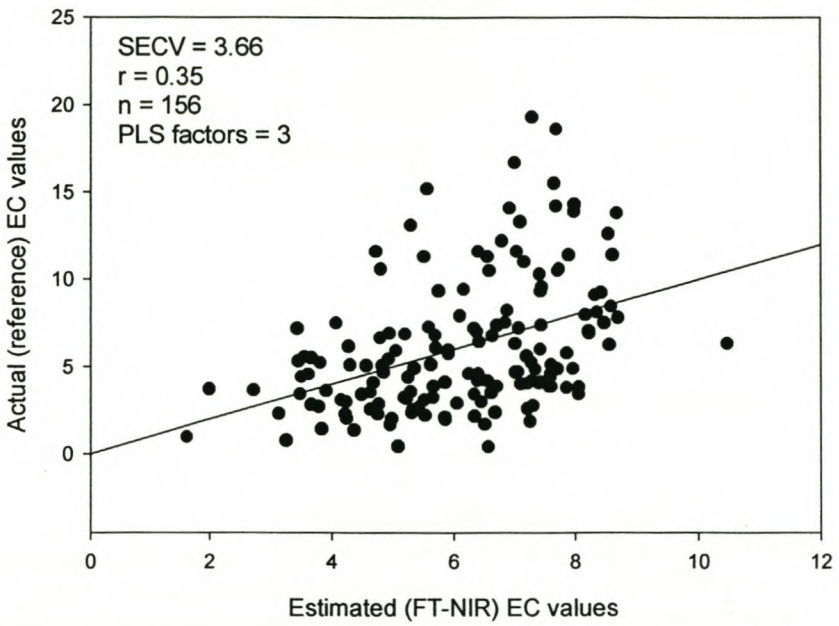


Figure 14. A plot of the estimated (FT-NIR) EC values versus actual (chemical analyses) EC values for the calibration model on wine samples.

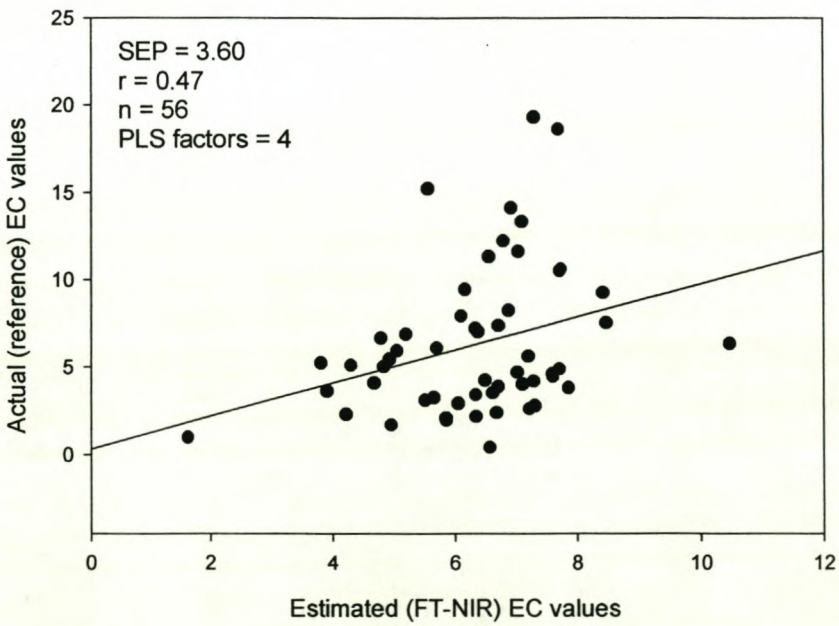
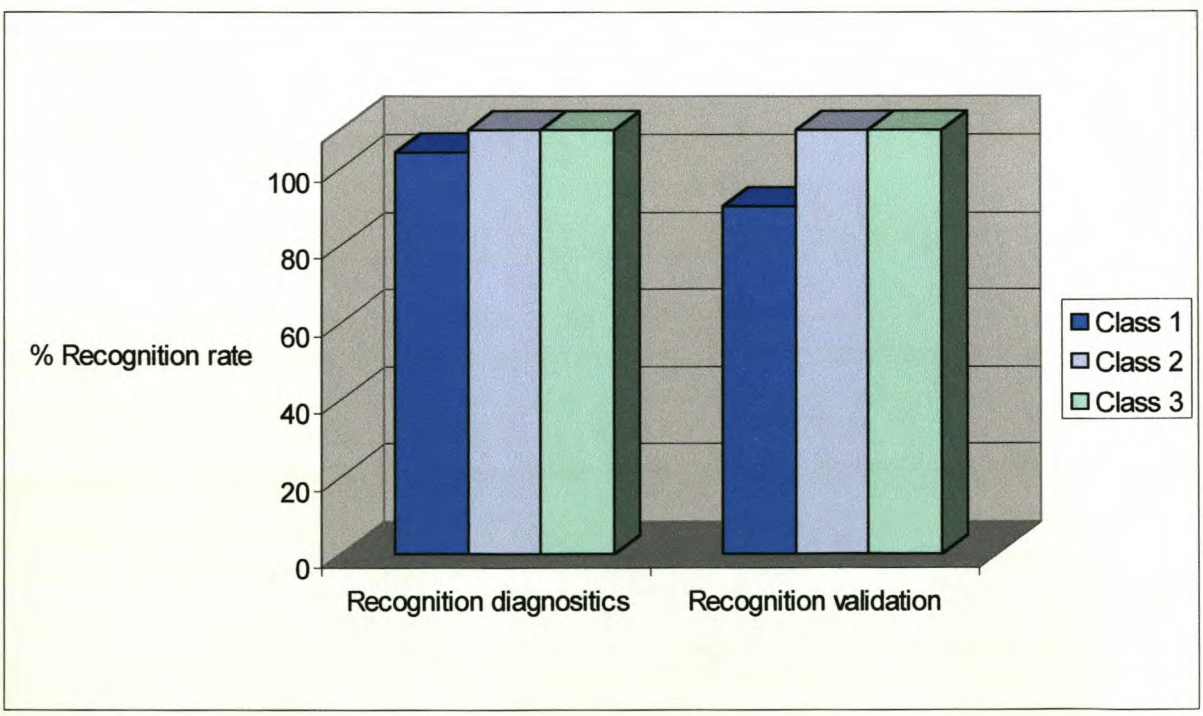


Figure 15. A plot of the estimated (FT-NIR) EC values versus actual (chemical analyses) EC values of the validation wine samples.

The recognition rate columns reported 94% for Class 1 and 100% for Classes 2 and 3 respectively, indicating that excellent separation of each class had been achieved. Summary of the verification diagnostic report is graphically shown in Figure 16. The three data sets were consequently tested using the validation procedure in the SIMCA analysis. This procedure validates the methods (data divided into classes) that have been built using test spectra that were removed from the data sets before the models were built, i.e. a validation of independent spectra. Again good results were obtained with recognition rates of 80% for Class 1 and 100% for Classes 2 and 3, respectively (Figure 16). The other 20% of Class 1 cannot necessarily be classified as belonging to Class 2 or 3, but indicate miss-classification. This could be due to the great similarity between the spectra as illustrated in Figure 17.

Figure 16. Graphic representation of SIMCA diagnostic and validation results.



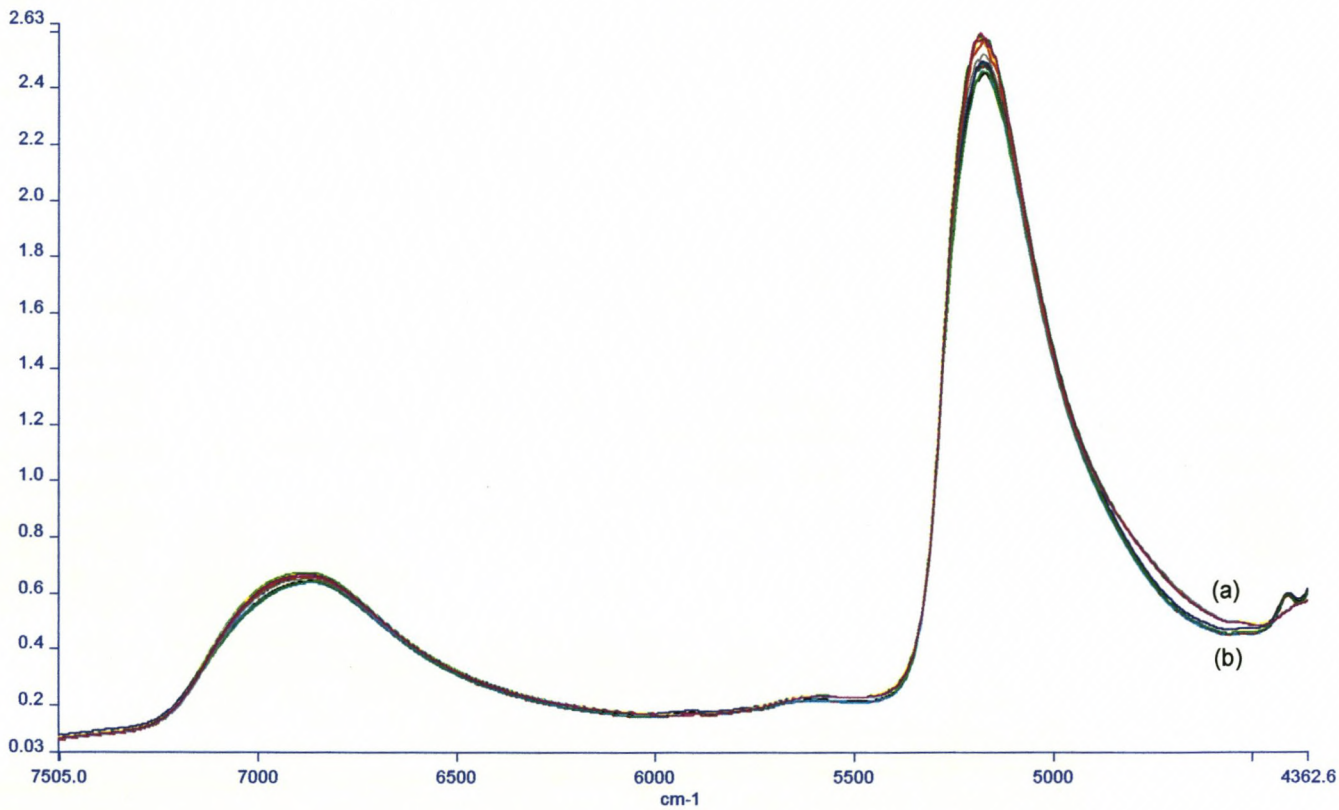


Figure 17. Spectra of must (a) and wine (b) samples.

Conclusion

This evaluation of the applicability of FT-NIR spectroscopy measurement of FAN and °Brix values, malic and lactic acid content and EC content in must and wine show considerable promise and may have immediate application in the wine industry. The conventional calibration method was tested, but inaccurate results obtained caused a shift in focal point to a classification chemometric method, SIMCA, which was applied with success. Due to the highly quantitative measurements required in a winery's laboratory, the conventional calibration method proved not accurate enough. A method such as SIMCA can discriminate between samples and has the potential to reduce the analytical times considerably in a range of measurements commonly used in determining the composition of the samples. In many processes it is only needed to know whether a specified cut-off point has been reached or not and can therefore replace expensive, time-consuming quantitative analytical methods, if not completely, at least to some extent.

References

Agenbach, W.A. (1978). A study of must nitrogen content in relation to incomplete fermentations, yeast production and fermentation. In: *Proceedings of the South African Society for Enology and Viticulture*. Pp. 66-87.

Amerine, M.A. & Ough, C.S. (1980). Wine and must analysis. John Wiley and Sons, New York.

Bisson, L.F. (1991). Influence of nitrogen on yeast and fermentation of grapes. In: *Proceedings of the International Symposium on Nitrogen in Grapes and Wine.* Pp. 78-89. Seattle, Washington, USA.

Butzke, C.E. & Bisson, L.F. (1997). Ethyl carbamate preventative action manual: Center for Food Safety and Applied Nutrition, Washington, DC, U.S. Food and Drug Administration.

Canas, B.J., Joe, F.L., Diachenko, G.W. & Burns, G. (1994). Determination of ethyl carbamate in alcoholic beverages and soy sauce by gas chromatography with mass selective detection: Collaborative study. *Journal of AOAC International*, **77**, 1530-1535.

Garcia-Jares, C.M. & Medina, B. (1997). Application of multivariate calibration to the simultaneous routine determination of ethanol, glycerol, fructose, glucose and residual sugars in botrytized-grape sweet wines by means of near-infrared reflectance spectroscopy. *Fresenius Journal of Analytical Chemistry*, **357**, 86-91.

Gishen, M. & Dambergs, B. (1998). Some preliminary trails in the application of scanning near infrared spectroscopy (NIRS) for determining the compositional quality of grapes, wine and spirits. *The Australian Grapegrower & Winemaker*, Annual Technical Issue 1998, 43-47.

Henschke, P.A. & Ough, C.S. (1991). Urea accumulation infermented grape juice. *American Journal of Enology and Viticulture*, **42**, 317-321.

Kunkee, R.E. (1991). Relationship between nitrogen content of must and sluggish fermentation. In: *Proceedings of the International Symposium on Nitrogen in Grapes and Wine.* Pp.148-155. Seattle, Washington, USA.

Monteiro, F.F., Trousdale, E.K. & Bisson, L.F. (1989). Ethyl carbamate formation in wine: use of radio actively labelled precursors to demonstrate the involvement of area. *American Journal of Enology and Viticulture*, **40**, 1-8.

Næs, T. & Isaksson, T. (1991). SEP or RMSEP, which is best? *NIR news*, **2**(4), 16.

Nielsen, J.C., Prahl, C. & Lonvaud-Funel, A. (1996). Malolactic fermentation in wine by direct inoculation with freeze-dried Leuconostoc oenos cultures. *American Journal of Enology and Viticulture*, **47**, 42-48.

Norris, K.H. (1989). Definition of NIRS Analysis. In: *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality* (edited by G.C. Marten, J.S. Shenk & F.E. Barton II). Pp. 6, 12. Agriculture Handbook No. 643. United Stated Department of Agriculture: Agriculture Research Service.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with practical applications in food and beverage analysis*, 2nd ed. Pp. 1-220. Essex, England: Longman Group UK Limited.

Ough, C.S., Stevens, D. & Almy, J. (1989). Preliminary comments on effects of grape vinyard nitrogen fertilization on the subsequent ethyl carbamate formation in wines. American Journal of Enolology and Viticulture, **40**, 198-202.

Perkin Elmer. (1997). *Spectrum QUANT+ User's Reference.* The Perkin Elmer Corporation. Beaconsfield, United Kingdom: Technical Publications.

Perkin Elmer. (1998a). A comparison of fiber optic and ICRA sampling for soft independent modeling by class analogy (SIMCA) classification of different celluloses. The Perkin Elmer Corporation. Beaconsfield, United Kingdom: FT-NIR Spectroscopy application notes.

Perkin Elmer. (1998b). Verification of tartrates used in the wine industry. The Perkin Elmer Corporation. Beaconsfield, United Kingdom: FT-NIR Spectroscopy application notes.

Schneider, A., Gerbi, V. & Redoglia, M. (1987). A rapid HPLC method for separation and determination of major organic acids in grape musts and wines. *American Journal of Enology and Viticulture*, **38**, 151-157.

Segal, M. (1988). Too many drinks spiked with urethane. *FDA Consumer*, http://vm.cfsan.fda.gov/~frf/fc0488ur.html. April, 1999.

Stevens, D.F. & Ough, C.S. (1993). Ethyl carbamate formation: Reaction of urea and citrulline with ethanol in wine under low to normal temperature conditions. American Journal of Enology and Viticulture, **44**, 309-321.

Volschenk, H., Viljoen, M., Grobler, J., Bauer, F., Lonvaud-Funel, A., Denayrolles, M., Subden, R.D. & Van Vuuren, H.J.J. (1997). Malolactic fermentation in grape musts by a genetically engineered strain of *Saccharomyces cerevisiae*. *American Journal of Enology and Viticilture*, **48**, 193-197.

Vos, P.J.A. (1977-1980). Determination of free amino-nitrogen in must or wine – outo analyser method. Wine and Spirit Research Report, Oude Meester Group Limited, R.I.T. Centre, Stellenbosch, South Africa.

Vos, P.J.A., Crous, E. & Swart, L. (1980). Fermentation and the optimal nitrogen balance of musts. *Wynboer*, March, 58-63.

Wehling, R.L. (1994). Infrared Spectrometry. In: Introduction to the chemical analysis of foods (edited by S.S. Nielsen). Pp. 344, 346-347, 349. Boston, MA: Jones and Bartlett Publishers.

Willard, H.H., Merritt Jr., L.L., Dean, J.A. & Settle Fr., F.A. (1988). Infrared Spectrometry. In: *Instrumental methods of analysis*, 7th ed. Pp. 302-303. California: Wadsworth Publishing Company.

Williams, P.C. (1991). Prediction of wheat kernel texture in whole grains by near-infrared transmittance. *Cereal Chemistry*, **68**, 112-114.

CHAPTER 4

NON-DESTRUCTIVE DETERMINATION OF THE TOTAL SOLUBLE SOLID CONTENT OF FRESH CLINGSTONE PEACHES USING FOURIER TRANSFORM NEAR INFRARED (FT-NIR) SPECTROSCOPY

Summary

The total soluble solid (TSS) content of fresh clingstone peaches are used as an indication of the maturity of the fruit for the canning industry. By determining this quality attribute, fresh peaches are graded on delivery and the farmers compensated accordingly. The sugar content of the peaches also need to be determined in order to ensure canned fruit with the correct cut-out °Brix value as dictated by the market. Fourier transform near infrared (FT-NIR) spectroscopy was used to perform quantitative TSS determinations on whole fresh clingstone peaches. Conventional soluble solid determinations requires maceration of the fruit prior to analysis, with the FT-NIR spectroscopic method being non-destructive which can be a suitable alternative method. Such a method would be of significant value to the South African peach canning industry as prediction of the soluble solid content of intact peaches is possible without sample preparation or destruction. Results obtained for the soluble solid content of clingstone peaches ($r = 0.96$, SEP = 0.55%) showed a good relationship between actual and predicted FT-NIR values.

Introduction

The canning of clingstone peaches have over the years developed into a large industry that produces approximately 4.1 million cartons of canned peaches annually in South Africa (W. Victor, Canning Fruit Producers' Association, personal communication). For an industry where such a large amount of fruit is

processed, an effective control system is essential to minimize losses. The total soluble solid (TSS) content (expressed and often referred to as °Brix) of fresh peaches in conjunction with firmness, are determined on reception at the canneries as an indication of fruit maturity. This is taken into account when producers are compensated for their fruit. Overripe fruit are more likely to disintegrate upon heat processing, while underripe fruit will necessitate the use of more sucrose for the preparation of the sugar syrup. The TSS content is an indication of the perceived sweetness. The cut-out °Brix (TSS) of the canned fruit is subjected to regulations set by the Department of Agriculture as well as specifications of respective clients. According to the Department of Agriculture, the minimum TSS is 18 °Brix for peaches on arrival with an allowed variations of 2 °Brix (G. Menzies, Langeberg Technical Services, personal communication). Present conventional methods for total soluble solid determinations require maceration and thus destruction of the samples prior to analysis. Fourier transform near infrared (FT-NIR) spectroscopy lends itself to a non-destructive means of determining TSS content. This instrumental method measures the chemical components of samples (Norris, 1989) and each of the major chemical components of a food sample has near infrared (NIR) absorption properties, which can be used to differentiate one component from the other.

Recent studies (Kawano *et al.*, 1992, 1993; Peiris *et al.*, 1997, 1998; Slaughter, 1995) show considerable work being done in regard to NIR prediction of metabolites occurring in peaches. Studies done in the United States of America (Peiris *et al.*, 1997, 1998) measured the soluble solids of whole fresh peaches using a refractometer to be used as NIR reference values. The multiple correlation coefficient (r) of the NIR calibrations ranged from 0.91 to 0.97 and the standard error of calibration (SEC) from 0.42% to 0.82%. Analysing Californian peaches, the optical absorption spectrum of each intact fruit was measured and then correlated with soluble solid content (r = 0.92, SEC = 0.87%), sucrose content (r = 0.87, SEC = 0.62%), sorbitol content (r = 0.88, SEC = 0.26%), sweetness index (r = 0.90, SEC = 0.70%) and chlorophyll A content (r = 0.97, SEC = 0.22%) (Slaughter, 1995). Similar studies done in Japan showed the

determination of TSS in intact peaches with an acceptable accuracy (r = 0.97, SEC = 0.48%) (Kawano *et al.*, 1992, 1993).

The soluble solid contents (°Brix), as measured with a refractometer, of a variety of other intact fruit were investigated by means of NIR spectroscopy by various researchers (Guthrie & Walsh, 1997; Kawano *et al.,* 1993; Moons *et al.*, 1997; Jordan *et al.*, 1997). Near infrared spectroscopy measurements of °Brix in intact pineapple and melons were executed with reasonable success (r = 0.72, SEP = 1.84%) (Guthrie *et al.*, 1998). Mangoes (Guthrie & Walsh, 1997), satsuma mandarins (Kawano *et al.,* 1993), apples (Moons *et al.*, 1997) and kiwifruit (Jordan *et al.*, 1997) are examples of other fruit that was subjected to quantitative internal quality assessment based on NIR spectroscopy.

Objective

The canning industry's products will be enhanced if a technique that can determine the quality attributes in whole fresh peaches at reception, without sample preparation or destruction, could be established. In this study, FT-NIR spectroscopy was applied to fresh clingstone peaches as a non-destructive analytical technique which can be used to measure its total soluble solid content. Such a determination is important for quality rating and maturity determinations but also to ensure that canned fruit have the correct cut-out °Brix value as dictated by the market and to determine the amount of sugar that needs to be added in order to produce a syrup of the required sweetness when canned.

Materials and methods

Samples

During the 1999 harvest season peaches from different fruit producing areas of the Western Cape Province of South Africa (mainly the Ashton district) were selected from different cultivars (Neethling, Sandvliet, Keisie, Katherina, Goudmyn, Oom Sarel, Malherbe, Woltemade and Kakamas) over a period of 6 weeks (middle January to the end of February). At the delivery depots the

peaches were packed into corrugated fiberboard cartons and transported to the ARC Infruitec-Nietvoorbij, Stellenbosch the day after harvesting. The fruit were placed in cold storage (-0.5°C) after numbering and FT-NIR analyses were performed the following day. The fruit were numbered individually and on both halves to allow matching of the constituent determinations with the corresponding FT-NIR spectra. The fruit were not washed, brushed or treated with any chemicals and the fruit's temperature was equilibrated to room temperature (21°C) prior to evaluation.

Spectral acquisition

The optical absorption spectrum was collected for each intact fruit using a Spectrum IdentiCheck™ 2.0 FT-NIR System (Perkin-Elmer) in reflectance mode. The absorption spectrum was collected by placing each fruit on the ICRA (IdentiCheck Reflectance Accessory) where monochromatic light interacted with the tissue of the fruit. Each fruit was analysed on both halves, 180° apart to obtain a better estimate of spatial variability, and the two scans were later averaged. Every absorption spectrum was the average of 16 individual optical scans in the wavelength range of 10 000 to 4000 cm^{-1} (1000 to 2500 nm) at intervals of 4 cm^{-1} to produced a total of 1501 data points per spectrum.

Sample analyses

Directly after the completion of the FT-NIR analyses, six peaches of each box were macerated for determination of TSS content. The total soluble solid content (expressed in °Brix) was determined with a Palette Digital refractometer. A sample of the macerated peach was filtrated through a No. 11 Whatman filter before the measurement. These results were used as reference data for the FT-NIR spectroscopic calibrations.

Data analyses

Calibrations were derived on second derivative spectra by means of partial least squares (PLS) regression to predict the total soluble solid content of the peach samples. The spectra were randomly divided into two sets: 47 samples (ca. 70%) were used for the calibration set and 23 samples (ca. 30%) for the independent validation set. Upon completion of the calibrations, the models were validated using independent sample sets.

The accuracy of the calibration to predict independent samples was expressed as the standard error of prediction, SEP of the bias-corrected residuals (p. 25). Alternatively the accuracy of the calibration was expressed as the root mean square error of prediction, RMSEP (p. 24). RMSEP is an estimate of the accuracy of the calibration against the reference method, and is calculated using an independent test set.

Results and discussion

The total soluble solid content of the calibration set ranged from 9 to 16.8 °Brix. When using 70 fruit samples with known TSS contents in the calibration data set, a correlation coefficient of 0.96 and a SECV of 0.58% (Table 1 & Figure 1) was found between FT-NIR and TSS content determined by chemical analyses. The accuracy of the calibration was then tested using the 23 spectra in the independent validation data set. The correlation for validation dropped slightly to 0.95 and the SEP to 0.55% with a bias of 0.006 (Table 1 & Figure 2).

$$SEL = \sqrt{\frac{\sum (y_1 - y_2)^2}{2n}} \qquad \ldots 1$$

Where: y_1 and y_2 = results of duplicate determinations

n = number of samples

A standard error of laboratory (SEL) of 0.20% was obtained for the laboratory results with a SEL of 0.5% being acceptable for °Brix (Van der Merwe, 1996) and therefore the FT-NIR method proved acceptable with a SEP of 0.55%.

Table 1. Summary of the statistical results obtained by the respective calibrations for total soluble solids (expressed as °Brix).

	Full cross validation		**Independent validation**
	Total soluble solids (%)		**Total soluble solids (%)**
Range	9 - 16.8	**Range**	9 - 16.8
Mean	13.51	**Mean**	13.45
SECV	0.58	**SEP**	0.55
RMSECV	1.38	**RMSEP**	0.57
BIAS	-0.05	**BIAS**	0.006
r	0.96	**r**	0.95
n (calibr.)	70	**n (calibr.)**	47
		n (indep.)	23
PLS factors	5	**PLS factors**	4
SD	1.7	**SD**	1.7

When the peach cultivar *shimizu Hakuto* from Okayama, Japan was used for similar studies, SEP values of 0.41% to 1.11% (Kawano *et al.*, 1995) were obtained for °Brix calibrations. Blake peaches from the United States of America showed the same tendency with SEP values for soluble solids calibrations between 0.62% and 1.52% (Peiris *et al.*, 1997), also proving suitable for rapid non-destructive determinations. In this study, SEP values of 0.55% for total soluble solid content of clingstone peaches were obtained, correlating well with previous studies.

Figure 3 illustrates the spectral differences between a selection of FT-NIR spectroscopic measurements of intact clingstone peaches.

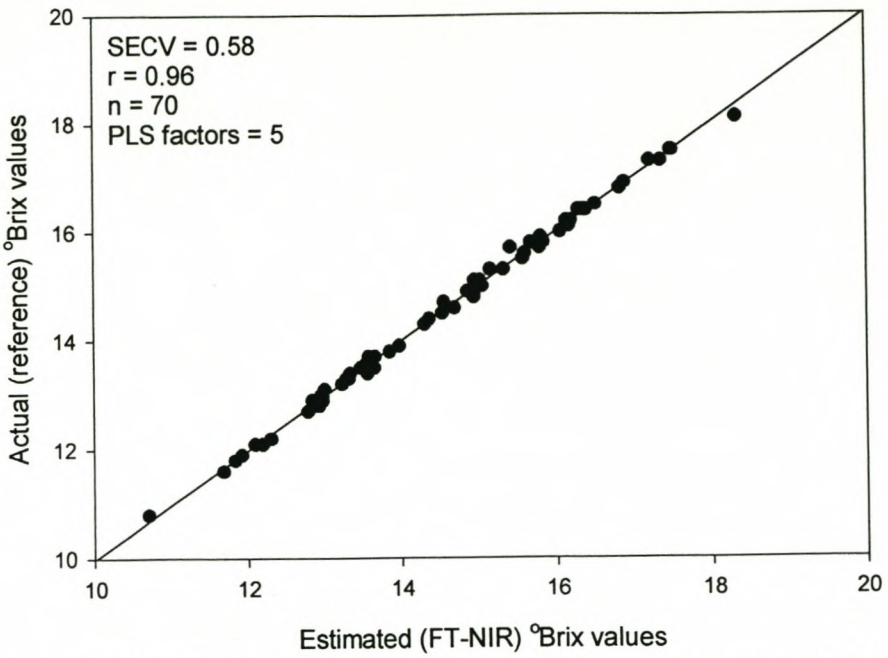


Figure 1. A plot of the estimated (FT-NIR) °Brix values versus the actual (physical analyses) °Brix values for the calibration model on peach samples.

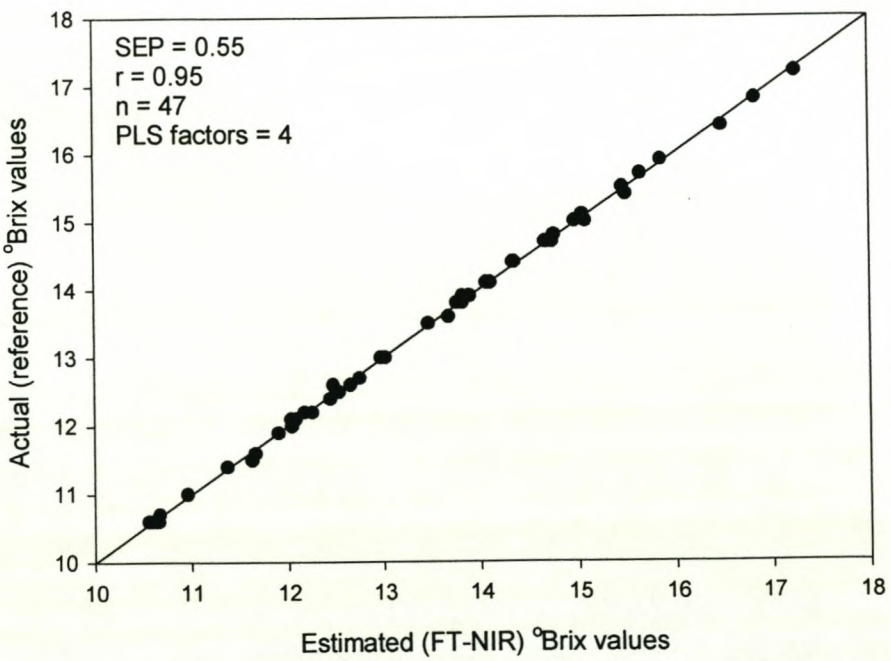


Figure 2. A plot of the estimated (FT-NIR) °Brix values versus the actual (physical analyses) °Brix values for the validation samples of fresh peaches.

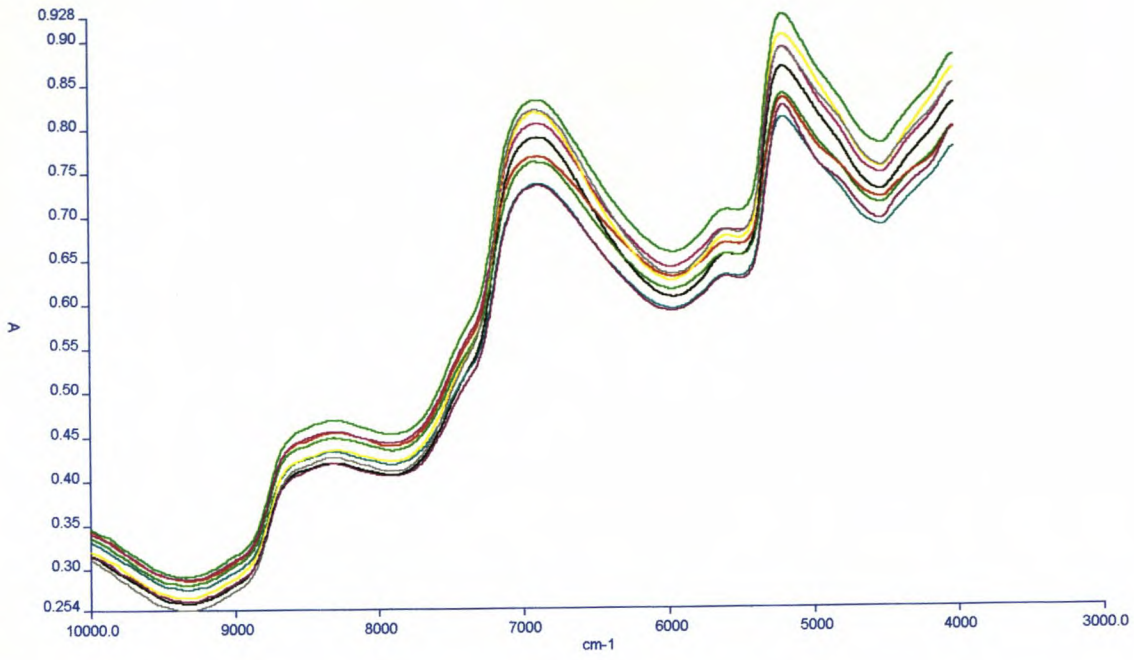


Figure 3. Spectra of fresh intact clingstone peaches.

Conclusion

Fourier transform near infrared spectroscopic measurements of total soluble solid content of whole fresh clingstone peaches proved acceptable to replace present methods for these determinations. This non-destructive method of analyses shows promising application possibilities for the canning industry as it can replace methods that presently require maceration prior to analysis.

References

Guthrie, J. & Walsh, K. (1997). Non-invasive assessment of pineapple and mango fruit quality using near infra-red spectroscopy. *Australian Journal of Experimental Agriculture*, **37**, 253-263.

Guthrie, J., Wedding, B. & Walsh, K. (1998). Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy*, **6**, 259-265.

Jordan, R.B., Kunnemeyer, R., Osborne, S.D. & Seelye, R.J. (1997). Harvest time prediction of eating time properties of kiwifruit using NIR transmission. In: *Proceedings from the Sensors for Nondestructive Testing International Conference and Tour*, Pp. 101-110. Orlando, Florida.

Kawano, S., Watanabe, H. & Iwamoto, M. (1992). Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in interactance mode. *Journal of the Japanese Society for Horticultural Science*, **61**, 445-451.

Kawano, S., Fujiwara, T. & Iwamoto, M. (1993). Nondestructive determination of sugar content in satsuma mandarin using near infrared (NIR) transmittance. *Journal of the Japanese Society for Horticultural Science*, **62**, 465-470.

Kawano, S., Abe, H. & Iwamoto, M. (1995). Development of a calibration equation with temperature compensation for determining the °Brix value in intact peaches. *Journal of Near Infrared Spectroscopy*, **3**, 211-218.

Moons, E., Dubois, A., Dardenne, P. & Sindic, M. (1997). Nondestructive visible and NIR spectroscopy for the determination of internal quality in apple. In: *Proceedings from the Sensors for Nondestructive Testing International Conference and Tour*, Pp. 122-132. Orlando, Florida.

Norris, K.H. (1989). Definition of NIRS Analysis. In: *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality* (edited by G.C. Marten, J.S. Shenk & F.E. Barton II). Pp. 6, 12. Agriculture Handbook No. 643. United Stated Department of Agriculture: Agriculture Research Service.

Peiris, K.H.S., Leffler, R.G., Dull, G.G. & Kays, S.J. (1997). Nondestructive determination of soluble solids content of peach by near infrared spectroscopy. In: *Proceedings from the Sensors for Nondestructive Testing International Conference and Tour*, Pp. 77-87. Orlando, Florida.

Peiris, K.H.S., Dull, G.G., Leffler, R.G. & Kays, S.J. (1998). Near-infrared spectrometric method for nondestructive determination of soluble solids content of peaches. *Journal of American Society of Horticultural Science*, **123**, 898-905.

Slaughter, D.C. (1995). Nondestructive determination of internal quality in peaches and nectarines. *American Society of Agricultural Engineers*, **38**, 617-623.

Van der Merwe, J.A. (1996). Maturity standards for deciduous fruit. In: *Integrated management of post-harvest quality* (edited by J.C. Combrink). Pp. 25-32. ARC Infruitec-Nietvoorbij.

CHAPTER 5

APPLYING FOURIER TRANSFORM NEAR INFRARED (FT-NIR) SPECTROSCOPY AND SOFT INDEPENDENT MODELLING BY CLASS ANALOGY (SIMCA) TO DETERMINE INTERNAL QUALITY OF WHOLE FRESH CLINGSTONE PEACHES USING SUBJECTIVE REFERENCE EVALUATIONS

Summary

The storage potential of fresh clingstone peaches can be successfully predicted with FT-NIR and SIMCA models using subjective internal quality evaluations. This will reduce losses, caused when storing peaches with poor storage quality, significantly. The peach canning industry is economically significant in South Africa with an annual production of ca. 150 000 tons of fresh peaches. Peak periods during the harvesting season necessitate storage of peaches for up to 3 weeks before canning. Approximately 5 - 15% of the peaches stored, disintegrate during canning due to loose skin, large stone cavities, soft flesh and rot. Maximum variability were incorporated into the calibration by using a sample set comprising of a large number of peaches (n = 2400) of different cultivars and from different localities. Results with recognition rates exceeding 80% were obtained in most cases.

Introduction

Peaches are one of the largest tree fruit crops in South Africa, with an annual production of between 125 000 to 150 000 tons (W. Victor, Canning Fruit Producers' Association, personal communication). Approximately 4.1 million cartons of canned peaches were produced during the 1999/2000 season resulting in producers' earnings of approximately R120 m. It is thus an economically significant industry of the Western Cape Province of South Africa

and relevant as a job creator in a country where unemployment is a nationwide concern.

Peaches are graded according to Proclamation R.2068 of Act 39 (1987) where the quality of peaches is determined based on minimum cross section, cultivar authenticity and shape, colour, ripeness, insect contamination and wounds and bruises by quality inspectors who randomly select at least 20 kg peaches per bulk bin. Fruit are also tested for pesticide residues to ensure compliance with regulations. Producers are then paid according to the grade the fruit have been assigned on delivery to the canning factory. Not all fruit are immediately canned, and can be subjected to cold storage for up to three weeks, especially during peak season. However, after cold storage, some of the consignments that were initially assigned a high quality grade, partially disintegrate upon destoning. This disintegration manifests itself in loosening of the skin from the flesh and a large stone cavity with part of the flesh removed with the stone. These losses must be kept to a minimum to improve profitability. According to the Canning Fruit Producers' Association of South Africa, losses due to storage of peaches are predicted between 5% and 15%, which costs the canneries between 6 and 15 million rand per annum (W. Victor, Canning Fruit Producers' Association, personal communication). These losses occur due to the lack in a suitable method of predicting storage potential of fresh peaches. The importance of accurate quality assessment of the fruit on reception is therefore very important to reduce losses. According to Kays (1999) the importance of non-destructive measurement of internal quality attributes of high moisture crops stems from the fact that quality evaluation (e.g. FT-NIR spectroscopy) of these crops is on the threshold of commercialisation.

Destructive testing is a common practice and such tests are conducted on a limited number of samples and are time-consuming, wasteful, and do not provide sufficient data for quality assessment of entire shipments of produce. An efficient, non-destructive method for evaluation of fruit would thus be useful. Peach canneries can increase their profits by canning a consistently high-quality product and by limiting their losses due to poor storage quality. In order to

establish such a consistency, the industry needs a means of quality control in its grading systems that can guarantee a consistently high level of quality. Near infrared spectroscopy offers the possibility of a rapid and non-destructive sensing technique for determining the quality of intact fruit.

The NIR spectroscopy method of analysis is an instrumental method for rapid and reproducible measurement of the chemical composition of samples (Norris, 1989). Each of the major chemical components of a sample has NIR absorption properties, which can be used to differentiate one component from the other. If the materials to be identified are spectroscopically dissimilar, it is often only necessary to use a simple distance measure such as a spectral difference. If the spectra are similar, it may be necessary to include more sophisticated techniques that take into consideration both the variability of the spectra of interest and the differences between the spectra (Perkin Elmer, 1998). The soft independent modelling by class analogy (SIMCA) technique provides such a method. SIMCA is a method that provides a set of parameters that characterise each class and are the basis for other quantities that describe the data.

Fruit and vegetables that lend themselves to quality assessment using NIR tend to have: (1) a quality parameter(s) that can be measured accurately; (2) individual product units that are of reasonably high value; and (3) quality attributes that cannot be ascertained readily by the customer (Kays, 1999). Previous NIR studies on the quality of fruit used °Brix and soluble solids as quality parameters (Kawano *et al.*, 1992; Kawano *et al.*, 1995; Peiris *et al.*, 1997).

The uniqueness of the present study lies in the fact that subjective quality evaluations were used to classify the peaches by use of FT-NIR spectroscopic methods. Based on parameters such as flesh on the stone, loose skin, slightly soft flesh, soft flesh, brown stone cavity and rot, the internal quality of the fruit and therefore the storage potential were evaluated.

Objective

Canneries are in need of a prediction model for evaluating the storage potential of canning peaches to reduce subsequent losses caused by storing. Peaches

with poor storage quality could then be canned as soon as possible after harvesting. The aim of this study was to develop a FT-NIR method based on subjective internal quality parameters to classify fresh clingstone peaches on reception at a cannery into classes, indicating their storage potential. A classification technique, SIMCA, was applied on the FT-NIR spectra of the peaches to discriminate between the samples.

Materials & methods

Samples

Different clingstone peach cultivars, selected from different fruit producing areas of the Western Cape Province of South Africa (Table 1) were packed into corrugated fiberboard cartons, at the delivery depots, and transported to the ARC Infruitec-Nietvoorbij, Stellenbosch the day after harvesting. The fruit were harvested from the end of January to the beginning of March 2000 over a period of 8 weeks. Every week 15 cartons, each containing 20 peaches, were collected for analyses and subjective quality evaluation, totaling 2400 peaches for the entire evaluation period. The fruit were numbered individually on both halves to allow matching of the subjective quality evaluation results with the corresponding FT-NIR spectra and kept at -0.5°C until FT-NIR analyses were performed on the following day. The fruit were not washed, brushed or treated with any chemicals and the fruit temperature was equilibrated to room temperature (21°C) prior to analyses.

Spectral acquisition

The optical absorption spectrum was collected on each intact fruit at the Department of Food Science, University of Stellenbosch, South Africa, using a Spectrum IdentiCheck™ 2.0 FT-NIR System (Perkin-Elmer) in reflection mode. The absorption spectrum was collected by placing each fruit on the ICRA (IdentiCheck Reflectance Accessory) where monochromatic light interacts with the fruit's tissue. Each fruit was measured on both cheeks, 180° apart to obtain a

Table 1. Summary of the different fruit producing areas, the peach cultivars and the number of cartons used for FT-NIR analyses and subjective evaluation.

	Area	Cultivar	Number of cartons
Week 1	Robertson	Neethling	2
	Ashton	Neethling	8
	Ceres	Neethling	2
	Wolseley	Neethling	2
	Ashton	Casey	1
Week 2	Montagu	Neethling	5
	Ashton	Neethling	8
	Ashton	Black	1
	Ashton	Malherbe	1
Week 3	Villiersdorp	Neethling	5
	Robertson	Neethling	5
	Bonnievale	Neethling	2
	Swellendam	Woltemade	3
Week 4	Robertson	Woltemade	3
	Robertson	Neethling	1
	Barrydale	Neethling	9
	Unknown	Unknown	2
Week 5	Ashton	Kakamas	2
	Robertson	Kakamas	1
	Unknown	Kakamas	3
	Ashton	Woltemade	3
	Tulbagh	Woltemade	1
	Ceres	Woltemade	3
	Unknown	Woltemade	2
Week 6	Ashton	Kakamas	4
	Robertson	Kakamas	9
	Bonnievale	Kakamas	1
	Unknown	Kakamas	1
Week 7	Tulbagh	Kakamas	2
	Donkerbos	Kakamas	2
	Robertson	Kakamas	1
	Breede River	Kakamas	5
	Koue Bokkeveld	Kakamas	2
	Unknown	Kakamas	3
Week 8	Ceres	Kakamas	15

better estimate of spatial variability and the two spectra were later averaged. Each absorption spectrum was the average of 16 individual optical scans in the wavelength range of 10 000 to 4000 cm^{-1} (1000 to 2500 nm) at intervals of 4 cm^{-1} to produced a total of 1501 data points per spectrum.

Sample evaluation

Directly after the completion of FT-NIR measurements, the fruit were stored for 3 weeks at -0.5°C. The fruit were subsequently lye peeled and destoned with a Filper destoner, similar to that used by the industry. Each peach was then subjectively evaluated for removal of excess flesh with the stone, loose skin, slightly soft flesh, soft flesh, brown stone cavity and rot. Using these subjective evaluations, the peaches were divided into classes based on the presence or absence of the above-mentioned parameters.

Data analyses

SIMCA models were built using spectral data and the classes based on subjective evaluation. SIMCA validations were performed to determine the accuracy of future classification of samples of unknown storage potential. Data collected for each individual week were analysed separately, followed by overall analyses for the total period of 8 weeks.

Results and discussion

In this study, no quantitative determinations were made as in the case of, for example, protein determinations in wheat where specific wavelengths represent certain functional groups. Linking subjective quality evaluations with FT-NIR spectroscopic data of peaches, relevant information might be found anywhere in the NIR wavelength range. Therefore, a very large sample set was used to incorporate as much variation as possible. Two thousand four hundred averaged spectra were obtained and divided into 5 initial classes based on the absence or presence of subjective evaluated defects (i.e. large and irregular stone cavity as

indication of flesh on the stone, loose skin, slightly soft flesh, soft flesh, brown stone cavity and rot). Class 1 contained all the samples with no defects after 3 weeks and thus peaches with best storage potential. Classes 2 to 5 contained the rest of the samples where some of the defects were present or not after 3 weeks storage. The different classes represented different combinations and quantities of defects (Table 2). From each class validation samples were selected to create independent validation sets to represent approximately a fifth of the total collection.

Table 2. Classes 1 to 5 representing different combinations of subjective quality defects.

	Flesh on the stone	**Loose skin**	**Slightly soft flesh**	**Soft flesh**	**Brown stone cavity and rot**
Class 1	not present	not present	not present	not present	not present
Class 2	not present	not present	present/not*	present/not	present/not
Class 3	present/not	present/not	not present	not present	not present
Class 4	present/not	present/not	present/not	present/not	present/not
Class 5	present	present	present	present	present

*The term present/not indicates the defect being present or not.

SIMCA diagnostics procedure validates every standard spectrum to ensure that the spectra from a single class fit that class (recognition) and that those selected from other classes selected are rejected (rejection). The recognition rate, also known as the sensitivity, is the number of spectra that are assigned to the class as a percentage of the number of spectra that should have been assigned to the class. The rejection rate, also known as the specificity, is the number of spectra that are rejected, thus not assigned to the class, as a percentage of the number of spectra that should have been rejected (Perkin Elmer, 1997).

The recognition rates reported were higher than 88%, indicating good classification (Table 3). Each data set was consequently validated. This procedure validates the methods (data divided into classes) that have been built using test spectra that were removed from the data sets before the models were

built; i.e. a validation of the independent spectra. Again good results were obtained, with the recognition rates reported being higher than 85%, indicating good separation (Table 3).

Table 3. Summary of the diagnostics and validation results for Classes 1 to 5 taken over 8 weeks.

	Diagnostics								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
Class1	109[a]/121[b]	112/123	92/96	80/85	19/19	37/37	20/20	63/64	
%	90[c]	91	96	94	100	100	100	98	**96**
Class2	24/24	87/98	63/65	91/101	63/74	145/155	53/57	57/58	
%	100	89	97	90	85	94	93	98	**93**
Class3	35/35	23/23	48/49	8/8	-*	-	8/8	36/37	
%	100	100	98	100	-	-	100	97	**99**
Class4	37/37	46/47	12/12	34/34	109/119	42/44	117/133	78/80	
%	100	98	100	100	92	95	88	98	**96**
Class5	27/28	10/10	18/18	10/10	17/18	-	23/23	7/7	
%	96	100	100	100	94	-	100	100	**99**
	Validation								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
Class1	21[a]/27[b]	24/24	22/23	17/21	4/5	9/9	4/4	14/15	
%	78[c]	100	96	81	80	100	100	93	**91**
Class2	5/5	18/19	14/16	22/25	14/17	33/38	14/14	14/14	
%	100	95	88	88	82	87	100	100	**92.5**
Class3	8/8	4/4	10/12	2/2	-	-	2/2	9/9	
%	100	100	83	100	-	-	100	100	**97**
Class4	6/9	9/9	2/2	6/8	23/29	10/11	27/33	16/19	
%	67	100	100	75	79	91	82	84	**85**
Class5	5/6	2/2	4/4	2/2	3/4	-	5/5	1/1	
%	83	100	100	100	75	-	100	100	**94**

a - Correctly recognized number of samples
b - Total number of samples
c - Percentage correctly recognized samples
* - No samples available for respective classes

Although high recognition rates were reported it was clear from the diagnostic results that some of the classes overlap spectroscopically. Consequently new classes were built, combining Classes 2, 3 and 4 into a single

class named Class B, and keeping Class 1 (now Class A) and Class 5 (now Class C). Class A represents all the samples with no subjective defects (those with the best storage potential), Class B contained the samples with one defect (irrespective of the kind of defect) and those in Class C having more than one defect (representing the peaches that have the worst storage potential) (Table 4). SIMCA diagnostic and validation procedures were applied to test the accuracy of the class separations and correctly classified on averaged 95% of the total diagnostics samples and 86% of the validation samples (Table 5).

Table 5. Summary of the diagnostics and validation results for the three quality classes, Classes A to C, taken over 8 weeks.

	Diagnostics								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
ClassA	109[a]/121[b]	91/99	92/96	80/85	19/19	37/37	20/20	63/64	
%	90[c]	92	96	94	100	100	100	98	**96**
ClassB	91/96	124/136	123/132	125/143	176/194	190/202	171/198	161/175	
%	95	91	93	87	91	94	86	92	**91**
ClassC	27/28	8/8	12/12	10/10	17/18	-*	23/23	7/7	
%	96	100	100	100	94	-	100	100	**98**
	Validation								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
ClassA	23[a]/28[b]	23/24	22/23	18/21	4/5	9/9	4/4	14/15	
%	82[c]	96	96	86	80	100	100	93	**92**
ClassB	17/21	24/32	29/32	29/35	26/46	31/49	37/49	29/42	
%	81	75	91	83	57	63	76	69	**74**
ClassC	5/6	2/2	2/2	2/2	3/4	-	5/5	1/1	
%	83	100	100	100	75	-	100	100	**94**

a - Correctly recognized number of samples
b - Total number of samples
c - Percentage correctly recognized samples
* - No samples available for respective classes

Table 4. Classes A to C representing different combinations of subjective quality defects.

	Flesh on the stone	Loose skin	Slightly soft flesh	Soft flesh	Brown stone cavity and rot
Class A	not present	not present	not present	not present	not present
Class B	present/not	present/not	present/not	present/not	present/not
Class C	present	present	present	present	present

*The term present/not indicates the defect being present or not.

Although it was possible to classify Classes A, B and C accurately and no spectroscopic overlap occurred, it might be ideal in practice to classify the peaches as only having acceptable storage potential or not. It was therefore decided to compile only two classes: one class (Class AA) containing all the samples with no defects after three weeks of storage and thus with acceptable storage potential; and another class (Class BB), representing all the samples evaluated as having any subjective defect indicating unacceptable storage potential (Table 6).

Table 6. Classe AA and Class BB representing the different combinations of subjective quality defects.

	Flesh on the stone	Loose skin	Slightly soft flesh	Soft flesh	Brown stone cavity and rot
Class AA	not present	not present	not present	not present	not present
Class BB	present/not	present/not	present/not	present/not	present/not

*The term present/not indicates the defect being present or not.

SIMCA diagnostic and validation procedures were applied to test the accuracy of the class separations and correctly classified on average 92 and 96% for the total diagnostics samples and 72 and 86% for the validation samples for Classes AA and BB, respectively (Table 7).

It is believed that certain subjective quality parameters are more relevant or directly linked to internal quality and storage potential than others, i.e. flesh on the stone resulting in a large and irregular stone cavity and loose skin. Based on

observations and experience, peaches with high nitrogen and moisture levels lead to characteristics such as flesh on the stone, large and irregular stone cavity and loose skin indicating peaches with inferior quality and thus a short storage potential (W. Victor, Canning Fruit Producers' Association, personal communication).

The respective Classes AA & BB were now streamlined for the canning industry by creating two new classes (Classes CC & DD) with newly developed SIMCA models. Class CC represents those samples with subjectively evaluated defects believed not to affect storage potential (slightly soft flesh, soft flesh, brown stone cavity and rot) as well as those with no defects, while Class DD represents those samples with subjectively evaluated defects such as flesh on the stone causing irregular and large cavities and the presence of loose skins (Table 8).

Table 7. Summary of the diagnostics and validation results taken over 8 weeks for Classes AA and BB.

	Diagnostics								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
Class AA	112[a]/117[b]	96/103	157/175	188/202	157/175	172/185	158/178	179/196	
%	96[c]	93	90	93	90	93	89	91	**92**
Class BB	120/125	123/137	26/26	30/31	33/35	54/56	44/45	43/45	
%	96	90	100	97	97	96	98	96	**96**
	Validation								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
Class AA	30[a]/48[b]	41/49	39/46	32/51	31/50	39/53	20/25	21/30	
%	63[c]	84	85	61	62	74	80	70	**72**
Class BB	8/11	10/11	11/12	8/8	7/7	5/6	28/34	20/29	
%	73	91	92	100	100	83	82	69	**86**

a - Correctly recognized number of samples
b - Total number of samples
c - Percentage correctly recognized samples

Table 8. The new classes, Class CC and Class DD, representing the different combinations of subjective quality defects.

	Flesh on the stone	Loose skin	Slightly soft flesh	Soft flesh	Brown stone cavity and rot
Class CC	not present	not present	present	present	present
Class DD	present	present	not present	not present	not present

SIMCA diagnostic procedures were applied to test the accuracy of the class separations and correctly classified on average 90% of the samples for Class CC and 92.5% for Class DD (Table 9). SIMCA validation procedures were also applied and correctly recognized an average of 80% and 90.5% of samples as belonging to Classes CC & DD, respectively (Table 9).

Table 9. Summary of the diagnostics and validation results taken over 8 weeks for Classes CC and DD.

	Diagnostics								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
Class CC	129[a]/142[b]	150/177	149/159	166/191	78/94	163/191	74/77	115/120	
%	91[c]	85	94	87	83	85	96	96	**90**
Class DD	99/110	62/67	80/84	48/52	150/165	52/54	147/169	116/121	
%	90	93	95	92	91	96	87	96	**92.5**
	Validation								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
Class CC	28[a]/39[b]	41/47	36/43	31/47	21/23	37/49	34/38	24/31	
%	72[c]	87	84	66	91	76	89	77	**80**
Class DD	17/20	13/13	15/15	12/12	31/34	9/10	19/21	19/28	
%	85	100	100	100	91	90	90	68	**90.5**

a - Correctly recognized number of samples
b - Total number of samples
c - Percentage correctly recognized samples
Class CC = Acceptable for storage
Class DD = Not acceptable for storage

In Table 9, the given diagnostics and validation results refer to the recognition rates obtained, indicating spectra with high sensitivity indicated by the high recognition rates. However, the rejection rates, which can also be obtained

from the SIMCA results, were not as high as the recognition rates (Table 10). This is expected due to the complexity of the intact tissue of a peach. When predicting the presence or absence of an analyte, recognition and rejection rates of similar accuracy is expected. In predicting internal quality, one does not search for evidence of an analyte itself, but acquire and analyse all the matrix data where all wavelengths might carry equal weight.

Table 10. Summary of the diagnostics and validation results taken over 8 weeks for Classes CC and DD, showing the low rejection rates.

	Diagnostics - rejection								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
Class CC	13[a]/142[b]	19/177	9/159	19/191	15/94	12/191	13/77	21/120	
%	9[c]	11	6	10	16	6	17	18	**12**
Class DD	29/110	13/67	11/84	15/52	59/165	11/54	27/169	15/121	
%	26	19	13	29	36	20	16	12	**21**
	Validation - rejection								
	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Average %**
Class CC	4[a]/39[b]	4/47	3/43	6/47	1/23	3/49	9/38	7/31	
%	39[c]	9	7	13	4	6	24	23	**16**
Class DD	8/20	3/13	3/15	3/12	8/34	2/10	0/21	8/28	
%	40	23	20	25	24	20	0	29	**27**

a - Correctly recognized number of samples
b - Total number of samples
c - Percentage correctly recognized samples

Figure 1 is a plot using the class distances and class index properties of two classes, Classes CC and DD calculated based on the degree of spectral simularity or differences. The critical distance for ClassCC was 1.16 and that of ClassDD 1.14 indicating a low degree of spectral differences.

The vertical and horizontal lines on Figure 1 represent the respective critical distances and divide the graph into quadrants. If the combined residual of the spectrum from a class is less than the critical distance, the spectrum is a member of the class. In quadrant 1 (indicated on Figure 1) the spectra with combined residual greater than the critical distance are found and thus not

belonging to the class. Figure 1 shows the low rejection rate, thus the lack of selectivity of the model caused by the low degree of spectral differences.

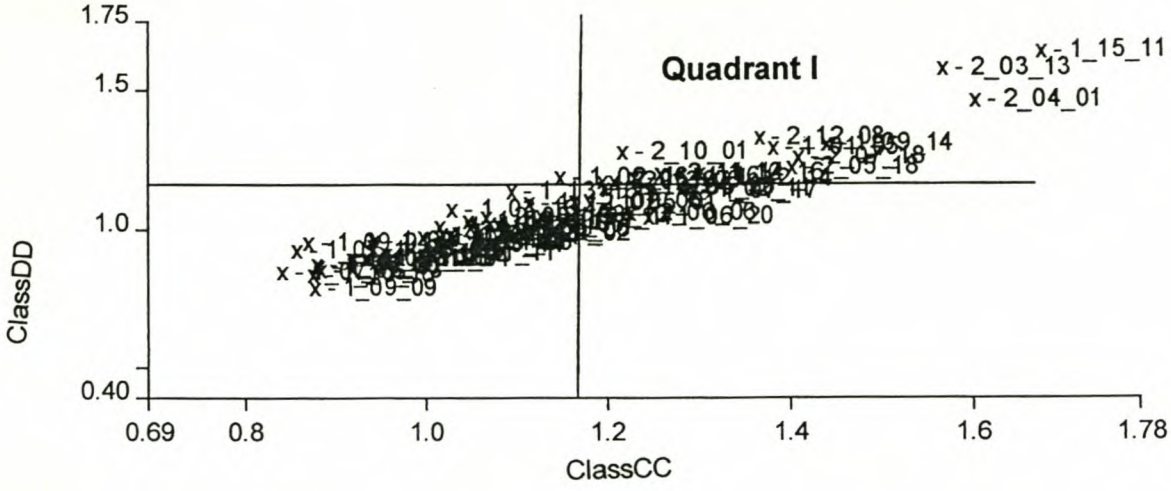


Figure 1. Graphical display of recognition, using the class distances which are influenced by the critical distances of the respective classes (1.14 for ClassCC and 1.16 for ClassDD).

Conclusion

Prediction of storage potential between fresh clingstone peaches on the basis of subjective internal quality parameter evaluations is possible using FT-NIR spectroscopy and SIMCA chemometric techniques.

The success of this unique application of calibrations with subjectively evaluated reference data has clearly been shown. This is an improvement in the prediction of internal quality and subsequent storage potential of fresh fruit compared to linking internal quality to analytical analyses. When predicting pure analytes, SIMCA results would be expected to be extremely accurate. However, when predicting internal quality of commodities such as whole fresh peaches, the entire complex data matrix is incorporated in the calculations, resulting in slightly less accurate predictions.

The current losses endured by the canning industry can be minimized by predicting the storage potential of the fruit using FT-NIR upon delivery to the cannery. Fruit with a poor storage potential can thus be processed as soon as possible. Non-destructive FT-NIR spectroscopy opens an exiting new field with numerous possibilities for the fresh fruit industries.

Reference

Kays, S.J. (1999). Non-destructive quality evaluation of intact, high-moisture products. *NIR news*, **10**, 12-17.

Kawano, S., Watanabe, H. & Iwamoto, M. (1992). Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in interactance mode. *Journal of Japanese Society of Horticultural Science*, **61**, 445-451.

Kawano, S., Abe, H. & Iwamoto, M. (1995). *Journal of Near Infrared Spectroscopy*, **3**, 211-218.

Norris, K.H. (1989). Definition of NIRS Analysis. In: *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality* (edited by G.C. Marten, J.S. Shenk & F.E. Barton II). Pp. 6, 12. Agriculture Handbook No. 643. United Stated Department of Agriculture: Agriculture Research Service.

Peiris, K.H.S., Leffler, R.G., Dull, G.G. & Kays, S.J. (1997). Nondestructive determination of soluble solids content of peach by near infrared spectroscopy. In: Proceedings from the Sensors for Nondestructive Testing International Conference and Tour, Pp. 77-87. Orlando, Florida.

Perkin Elmer. (1997). *Spectrum QUANT+ User's Reference.* The Perkin Elmer Corporation. Beaconsfield, United Kingdom: Technical Publications.

Perkin Elmer. (1998). A comparison of fiber optic and ICRA sampling for Soft Independent Modeling by Class Analogy (SIMCA) classification of different celluloses. The Perkin Elmer Corporation. Beaconsfield, United Kingdom: Application notes.

CHAPTER 6

DETERMINATION OF SO_2 AND MOISTURE CONTENTS OF GOLDEN SULTANAS BY MEANS OF FOURIER TRANSFORM NEAR INFRARED (FT-NIR) SPECTROSCOPY

Summary

Fourier transform near infrared (FT-NIR) spetroscopy was investigated as a rapid and accurate means of determining the SO_2 and moisture contents of golden sultanas. Rapid assessment of SO_2 and moisture contents is needed at delivery depots as a large number of consignments must be evaluated on arrival and the current time-consuming conventional methods tend to be a hindrance. During processing rapid analyses of moisture and SO_2 contents would facilitate prompt decision making when required. High positive correlations was found between the measured values and those predicted by FT-NIR spectroscopy for SO_2 ($r = 0.99$, SEP = 32 $mg.kg^{-1}$) and moisture ($r = 0.99$, SEP = 0.045%) contents.

Introduction

Golden sultanas are produced by sulphuring of Sultanina grapes before drying and are mainly used for baking purposes. The golden sultana industry of South Africa plays an important role in the exporting market of dried fruit. The export value of these sundried golden sultanas is approximately R10 000 per ton (D. Smit, Dried Fruit Technical Services, personal communication). During the 1999 harvest season 8500 tons of golden sultanas were produced, resulting in earnings of R85 million, stressing the economical significance of this industry.

Sulphuring is essential for development of the desired colour and one of the important quality factors when grading sultanas. The other two important quality factors are the SO_2 and moisture contents of the sultanas. On delivery to

the factory, golden sultanas may not exceed a SO_2 content of 1500 $mg.kg^{-1}$ and a moisture content of 15% (Anonymous, 1996).

The moisture content of sultanas is determined upon delivery to the processing plant as a means of quality assessment. SO_2 and again moisture contents are determined during processing and the more accurate, however, much more time-consuming Monier-Williams method (60 minutes to complete) is used for SO_2 determinations (J. Marais, S.A. Dried Fruit Co-operative LTD, personal communication). The conventional vacuum oven method used for moisture determinations, is also time-consuming, needs maceration of the sample and cannot be used for large samples. The electronic resistance method is sometimes used on delivery for larger samples, but it also needs maceration prior to analysis (D. Smit, Dried Fruit Technical Services, personal communication).

The large number of consignments that must be graded and analysed upon arrival, and the need for rapid decisionmaking during processing, make replacement of time-consuming analytical methods essential.

Objective

The aim of this study was to develop a rapid and accurate FT-NIR spectroscopic method of analyses to replace the present time-consuming methods when determining SO_2 and moisture contents of macerated golden sultanas.

Materials and methods

Samples

Golden sultana samples (n=98) from the Northern Cape Province of South Africa were collected by the South African Dried Fruit Co-operative LTD (SAD), Upington and sent to the ARC Infruitec-Nietvoorbij, Stellenbosch for analyses. The samples were cold stored at 5°C to reduce SO_2 losses until being analysed. On the day of analyses, the temperature of the samples were equilibrated to

room temperature (21°C), before maceration. Directly after maceration, the samples were sub-divided into 2 parts: one part for FT-NIR spectroscopic analyses; and the other for SO_2 and moisture determinations, respectively.

Chemical analyses

The Tecator method (Anonymous, 1978) was used to determine the SO_2 content of each macerated golden sultana sample in duplicate. Duplicate moisture analyses were performed using a vacuum oven drying method. The samples (5 g) were weighed in identical moisture dishes and dried under vacuum for 16 hours at 70°C, whereafter the mass loss was determined.

Spectral analyses

The optical absorption spectra were collected for each macerated sultana sample by placing a glass dish filled with sample on the rotating ICRA (IdentiCheck solids sampling accessory) attached to the FT-NIR spectrophotometer (Perkin-Elmer Spectrum IdentiCheck™ 2.0 System). These spectra were recorded over the wavelength range of 10 000 to 4000 cm^{-1} at 4 cm^{-1} intervals. The rotating ICRA allows a number of scans to be taken at different areas on the sample and averages those scans.

Calibrations developed

Data manipulation were performed using QUANT+™ 4.1 software (Perkin Elmer). The spectra were randomly divided into two sets: 66 samples (ca. 70%) were used for the calibration set and 32 samples (ca. 30%) for the validation set. Sulphur dioxide and moisture calibrations were derived on spectra on which multiplicative scatter correction (MSC) and second derivative processing have been applied. Each model was also validated on an independent validation sample set. Partial least squares (PLS) regression (Haaland & Thomas, 1988), a full spectrum data reduction calibration method, that utilize absorbance data from each measured wavelength of the spectrum, was used to derive calibrations.

Cross validation (Osborne *et al.*, 1993) was performed to construct PLS factors for the original spectral data by removing one sample from the calibration set and calibrate the remaining samples followed by predicting the sample that has been removed. This procedure was repeated until all samples were removed from the calibration set and predicted once.

Results and discussion

A very strong correlation existed between the FT-NIR spectroscopic predictions of SO_2 in the golden sultana samples and that of the actual SO_2 values. When the calibration model was validated by means of cross validation, the variance between the samples (SECV) was 7.76 $mg.kg^{-1}$ and the variance within the samples (BIAS) was 3.21 $mg.kg^{-1}$ with a r of 0.99 (Table 1 & Figure 1). When performing the independent validation, a correlation coefficient of 0.99 and an excellent SEP of 32 $mg.kg^{-1}$ were obtained (Table 1 & Figure 2). Due to the possibility of overfitting when 6 principal components (PC's) are used with 66 calibration samples the calibration was repeated with 4 PC's (Figure.5) and still an acceptable SEP of 88.87 $mg.kg^{-1}$ and r of 0.98 were obtained (Table 2).

The standard error accepted in the laboratory (SEL) for SO_2 determinations using the Tecator method is 100 $mg.kg^{-1}$ SO_2. Although the NIR method is seldom more accurate than the reference method, in this case an excellent accuracy has been obtained. Another indication of the efficiency of a calibration, is the standard deviation of reference data (RPD) (Williams, 1991) which is the standard deviation (SD) between the samples in the set divided by the SEP or SECV of that respective model. A RPD value above 8.1 can be interpreted as an excellent prediction with any application possibility (P.C. Williams, Canadian Grain Commission, personal communication). In the case of the SO_2 prediction, the RPD value for the validation set was 10.18 confirming the excellent calibration performance.

The moisture calibration model resulted in a good correlation between the FT-NIR spectroscopic predictions and the actual values obtained by the vacuum oven drying method with an r and SECV of 0.99 & 0.023% (Figure 3 & Table 1)

and a r and SEP of 0.99 & 0.045% (Figure 4 & Table 1) for the calibration and validation sets, respectively. Again the calibration was repeated by reducing the PC's to 4 to prevent the possibility of overfitting (Figure 6) and a r and SEP of 0.99 & 0.135% were obtained (Table 2). The results indicated that FT-NIR spetroscopy can predict sultana moisture extremely accurately. This is expected as NIR has been reported as a very good predictor of moisture for a wide range of food and feed materials (Williams & Norris, 1987).

Table 1. Summary of the validated calibrations' statistical results of the macerated golden sultana samples.

	Full cross validation			**Independent validation**	
	SO_2 ($mg.kg^{-1}$)	**Moisture**(%)		**SO_2** ($mg.kg^{-1}$)	**Moisture** (%)
Range	403.48 - 1890.54	9.93 - 16.51	**Range**	403.48 - 1890.54	9.93 - 16.51
Mean	830.60	12.30	**Mean**	830.60	12.30
SECV	7.78	0.02	**SEP**	32	0.045
BIAS	-3.21	0.02	**BIAS**	49.46	0.09
r	0.999	0.999	**r**	0.999	0.999
n	98	98	**n (calibr.)**	66	66
			n (indep.)	32	32
PLS factors	9	8	**PLS factors**	6	6
SD	326.01	1.42	**SD**	326.01	1.42
RPD	41.93	60.75	**RPD**	10.18	31.5

The SEL and SEP for the moisture predictions are 0.5% and 0.05% respectively, indicating the FT-NIR method as a highly efficient alternative method. The calibration for the moisture content showed once again to be very efficient in terms of the RPD statistics with a value of 31.5 for the validation set (P.C. Williams, Canadian Grain Commission, personal communication). Huxsoll *et al.* (1995) predicted the moisture content in raisins from Thomson seadless grapes with an accuracy of 0.88%.

FT-NIR spectra of a selection of macerated golden sultana samples, are illustrated in Figure 7.

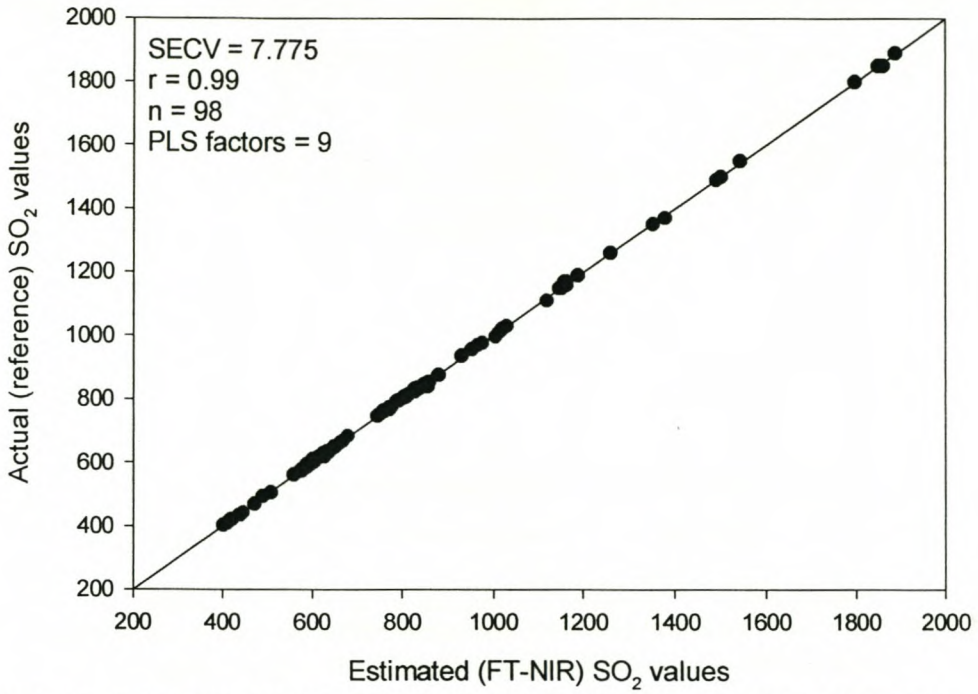


Figure 1. A plot of the estimated (FT-NIR) SO_2 values versus actual (chemical analyses) SO_2 values for the calibration model on sultana samples.

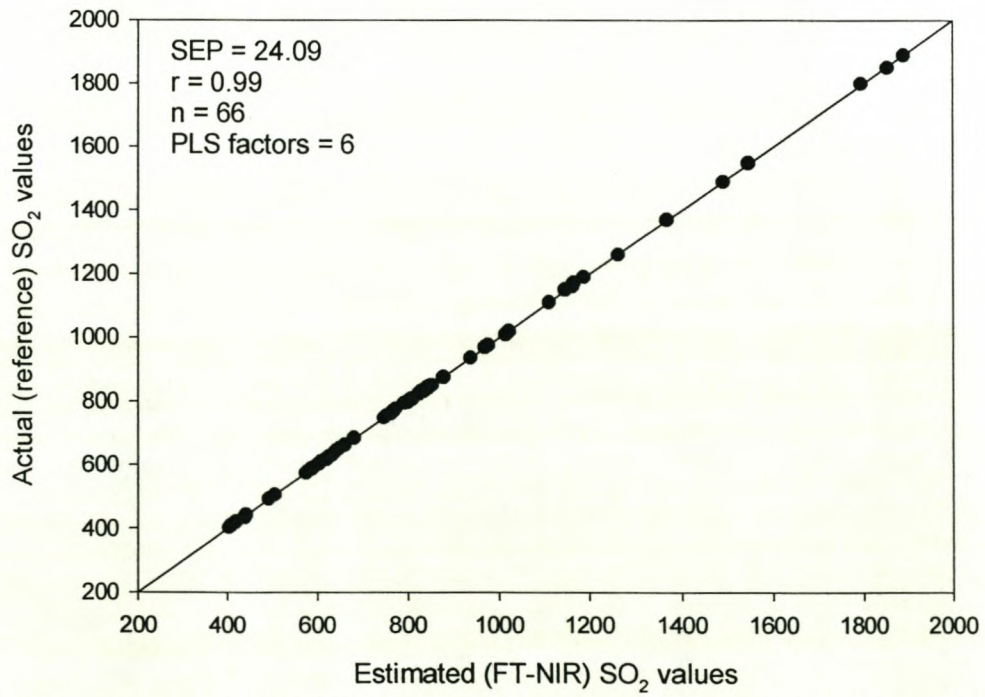


Figure 2. A plot of the estimated (FT-NIR) SO_2 values versus actual (chemical analyses) SO_2 values for the validation sultana samples.

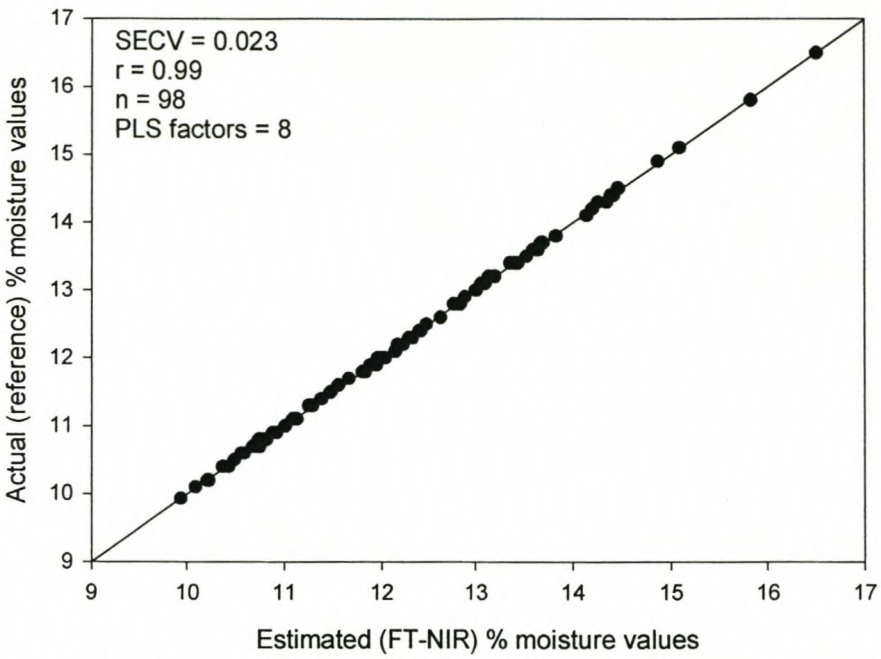


Figure 3. A plot of the estimated (FT-NIR) % moisture values versus actual (chemical analyses) % moisture values for the calibration model of sultana samples.

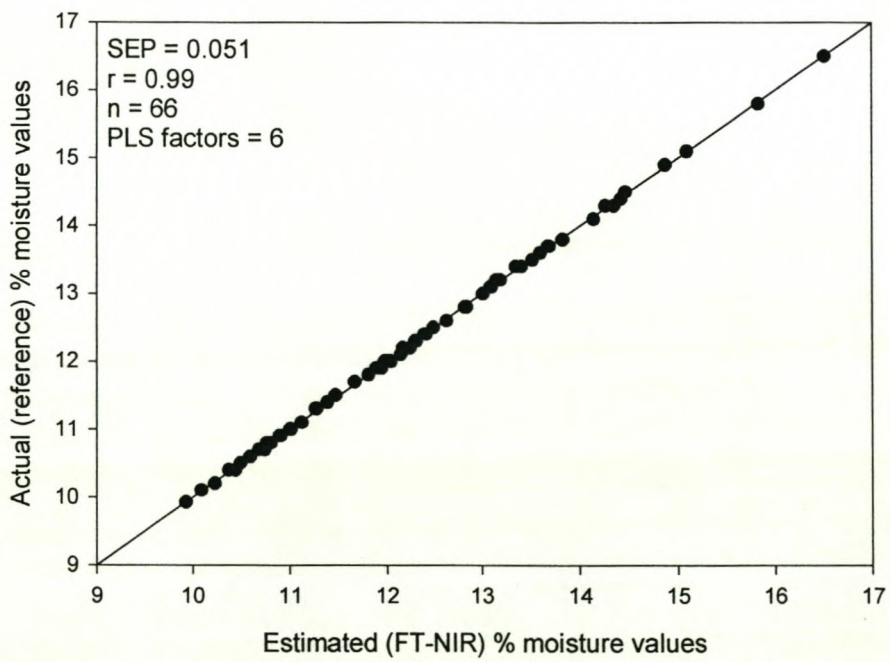


Figure 4. A plot of the estimated (FT-NIR) % moisture values versus actual (chemical analyses) % moisture values for the validation sultana samples.

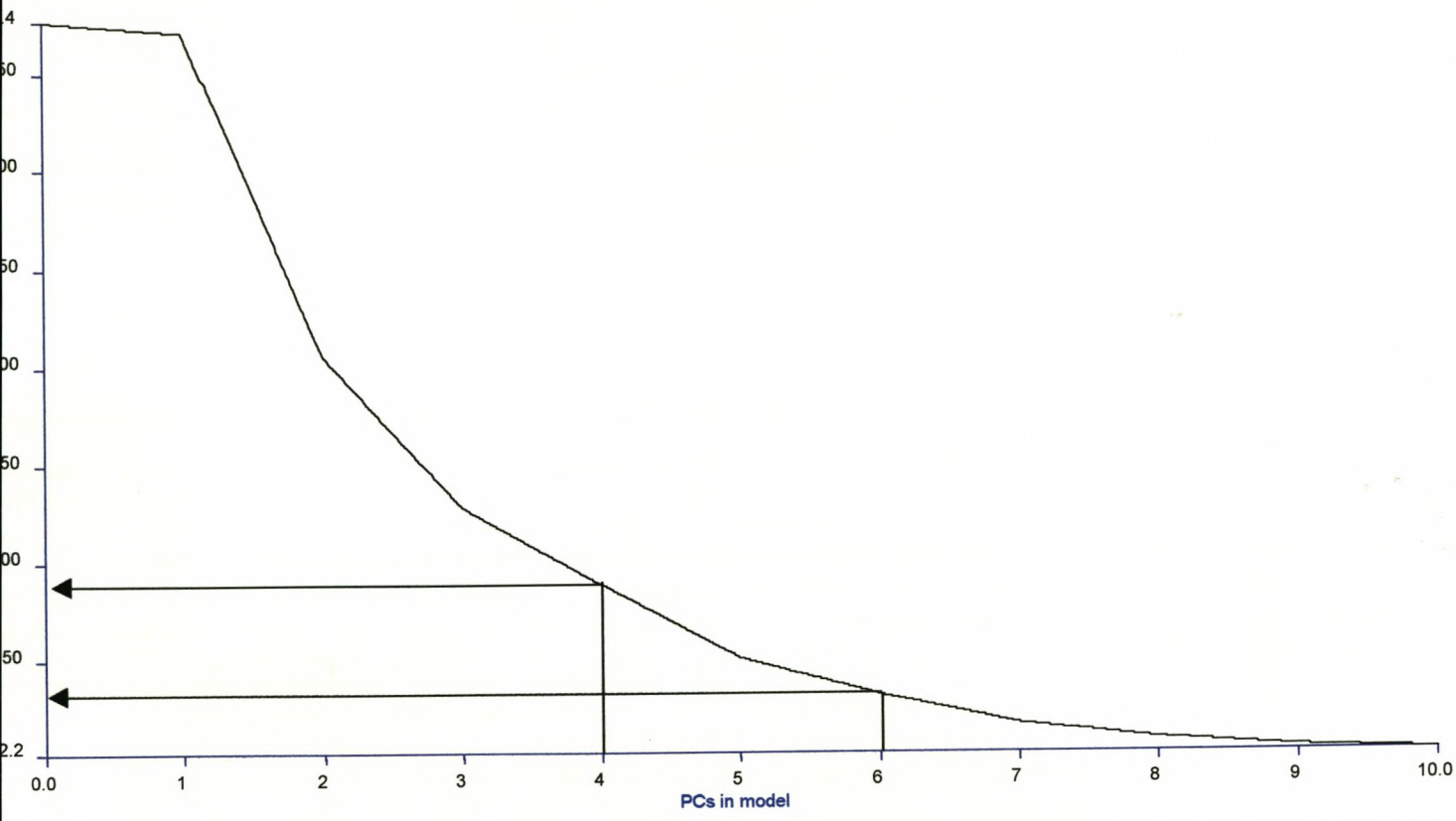


Figure 5. The residual variance plot of the SEP values versus the numbers of principal components for the SO_2 calibration model on macerated golden sultana samples.

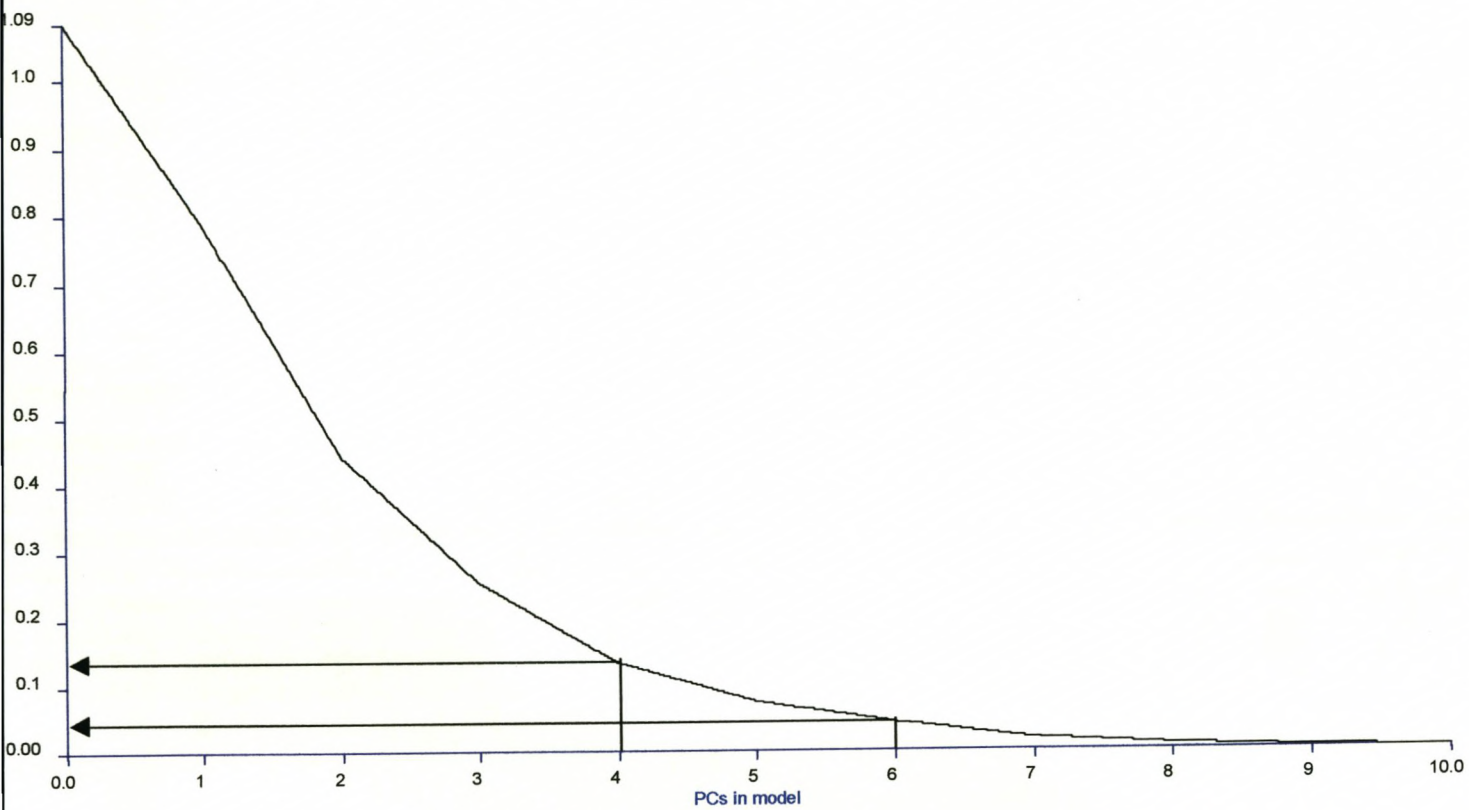


Figure 6. The residual variance plot of the SEP values versus the number of principal components for the moisture calibration model on macerated golden sultana samples.

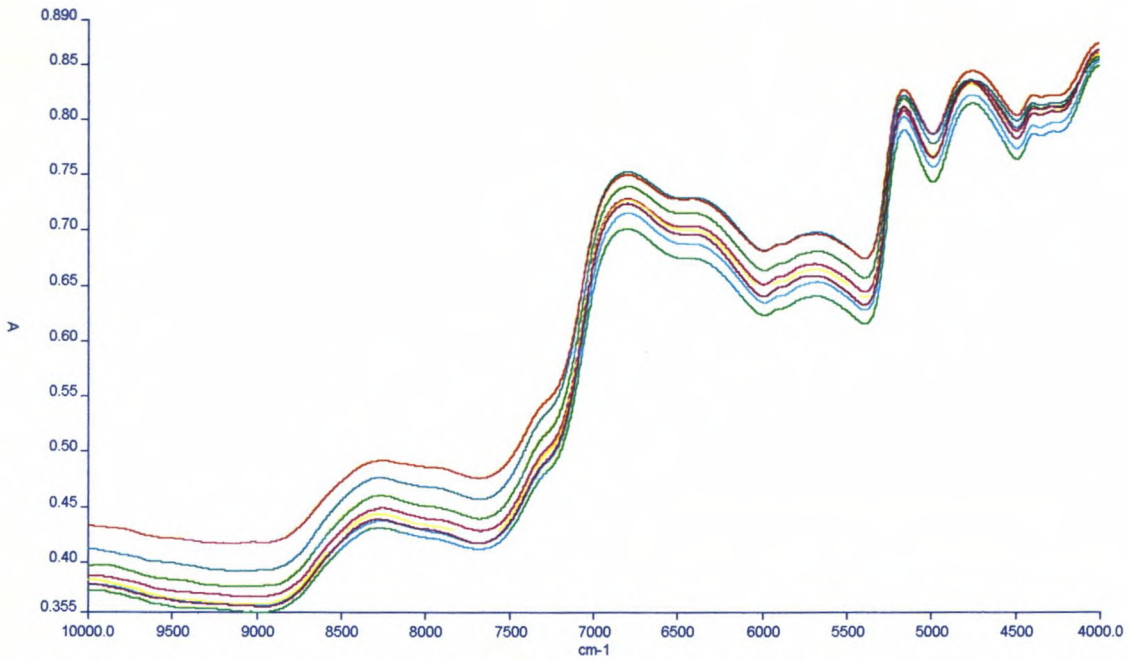


Figure 7. FT-NIR spectra of macerated golden sultana samples.

Table 2. Summary of the independent validated calibration's results of the macerated sultana samples when different principal components were used.

	Independent validation			
	SO_2 ($mg.kg^{-1}$)	**SO_2** ($mg.kg^{-1}$)	**Moisture** (%)	**Moisture** (%)
RMSEP	360.4	361.6	0.696	0.7
SEP	32	88.87	0.045	0.014
BIAS	49.46	47.65	0.092	0.115
RMSEP	0.99	0.78	0.99	0.99
n (calibr.)	66	66	66	66
PLS factors	6	4	6	4

Conclusion

FT-NIR spetroscopy proved to be suitable as a reliable method for predicting the SO_2 and moisture contents of macerated golden sultanas by scanning the samples in a glass dish on a rotating ICRA. FT-NIR spectroscopy can replace conventional time-consuming methods successfully. In both cases the predictions were extremely accurate and could this method be established in industry. This method also has the potential to be non-destructive as NIR spetroscopy does not require sample preparation when using suitable accessories and should be investigated in the laboratory.

References

Anonymous. (1978). The determination of sulphur dioxide. *Tecator Application Note*, June, 1-4.

Anlonymous. (1996). Regulations and guidence for ingrading of dried fruit. S.A. Dried Fruit Co-operative LTD. Pp. 10-11.

Haaland, D.M. & Thomas, E.V. (1988). Partial least-squares methods for spectral analyses. *Analytical Chemistry*, **60**, 1193-1202.

Huxsoll, C.C., Bolin, H.R. & Mackey, B.E. (1995). Near infrared analysis potential for grading raisin quality and moisture, *Journal of Food Science*, **60**, 176-180.

Osborne, B.G. & Fearn, T. (1986). *Near Infrared Spetroscopy in Food Analysis*. Pp. 136-158. Essex, England: Longman Scientific and Technical, New York: John Wiley & Sons, Inc.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Practical Applications in Food and Beverage Analysis*, 2nd ed. Pp. 1-220. Essex, England: Longman Group UK Limited.

Williams, P. & Norris, K. (1987). *Near-Infrared Technology in the Agricultural and Food Industries*. Pp. 1781-1823. American Association of Cereal Chemists, St. Paul, MN.

Williams, P.C. (1991). Prediction of wheat kernel texture in whole grains by near-infrared transmittance. *Cereal Chemistry*, **68**, 112-114.

CHAPTER 7

GENERAL DISCUSSION AND CONCLUSION

A clear need exists for rapid non-destructive methods that can be used for routine analyses, whether it is the vibrant wine industry or the established peach canning and dried fruit industries of South Africa. In all of these industries various routine analysis are carried out as an essential part of process or quality control, and Fourier transform near infrared (FT-NIR) spectroscopy has been investigated as an alternative method. This study is the first of its kind in South Africa as no other FT-NIR work has before been done on wine, fruit and dried fruit in this country.

The diverse wine industry consists of very small to large wineries. In the larger wineries a huge number of routine analysis are executed daily to monitor the status of processes taking place during wine-making and to ensure that wines comply with the necessary export and local regulatory demands. A number of analytical instruments and processes that provide highly quantitative results are used, however, in many cases it is only needed to know whether a process has reached a certain stage or not. Methods that can streamline present time-consuming routine work were investigated in the present study for the wine industry. Such methods should allow the screening of processes where it is only necessary to know whether a certain cut-off points has been reached or not. This will consequently reduce the number of samples to be analysed quantitative to a great extent as only the samples close to the cut-off points, will have to be subjected to more sophisticated analyses.

FT-NIR spectroscopy and SIMCA were applied to develop discriminative models for the determinations of free amino nitrogen (FAN) values, the status of the malolactic fermentation (MLF) and the level of ethyl carbamate (EC) present that have to be executed during wine-making. Good classifications were obtained proving the SIMCA classification method to be useful. Small wineries will probably not invest in the NIR technology, but larger wine producers and

independent wine laboratories will surely benefit from this method as they can save time and money when the number of quantitative analyses are cut to a minimum. Implications can arise when wineries that export local wine to countries where the EC content of the wines are restricted by legislation, have no knowledge of the EC content of the wines they are exporting. It is foreseen that in the near future all export wines will have to undergo EC content analyses and a rapid screening method will provide a welcome solution.

When FT-NIR spectroscopy was applied on fresh clingstone peaches, the objective was to develop among others a method to replace the present refractometer method that requires maceration of the sample prior to analyses to determine the total soluble solid (TSS) content (expressed as °Brix) of the peaches on reception to the cannery. A method that fulfil this objective was developed in the form of a successful FT-NIR calibration of TSS content. No sample preparation is required (e.g. maceration) and the intact peaches can simply be subjected to FT-NIR analysis with the accurately predicted TSS values as the result. The possibility to apply NIR on-line, will provide a method to sort the peaches according to quality parameters and the cut-out amount of sugar used by the industry.

On arrival at the canneries, peaches need to be classified as being of sufficient quality to be stored for up to 3 weeks. If not, the peaches have to be canned as soon as possible to reduce losses. Such a classification method was requested as the present method of determination is subjective and depends on the experience of the inspector. FT-NIR spectroscopy was applied and SIMCA models were created, using subjectively evaluated reference data to develop a classification method that met their needs. When applying FT-NIR, there was found to be an 80% chance to predict correctly whether the peaches need to be canned immediately or whether they can be stored for up to 3 weeks.

Future research work being planned for this project, is the development of a field model NIR spectrophotometer. Testing of the equipment on location will give an indication of the success of a model created under laboratory conditions. If positive results are obtained from the field model study, the next step should be

testing an on-line discriminative method for classification purposes. This technique has the potential to save the industry immense amounts of money.

Another unique study for South Africa was the application of FT-NIR spectroscopy on golden sultanas. Time-consuming analytical processes are presently being used for SO_2 and moisture content determinations and the possibility of the rapid NIR method replacing these conventional methods had to be investigated. Great success was achieved with SO_2 and moisture content determinations and excellent results were obtained. These open exciting new application possibilities for the dried fruit industry. The application of FT-NIR on whole sultanas should be looked into to eliminate the need for maceration of the fruit and the success of such a model will not only produce a rapid method of analyses, but also one that is entirely non-destructive.

Expensive, time-consuming, sophisticated analytical methods are being used to a great extent in the wine, fruit and dried fruit industries, monitoring quality parameters in raw materials and during processing. In this study FT-NIR has been shown to be an alternative method, if not completely at least to some extent to replace the present methods of determination. Additional benefits of FT-NIR are multiple determinations on the same samples and on-line applications possibilities.